



Universidade Federal do Rio de Janeiro - UFRJ
Centro de Ciências Matemáticas e da Natureza
Instituto de Geociências
Departamento de Meteorologia

**PREVISÃO DE TEMPO SEVERO CONVECTIVO UTILIZANDO
INTELIGÊNCIA COMPUTACIONAL**

JIMMY NOGUEIRA DE CASTRO

Rio de Janeiro, RJ
Outubro de 2021



PREVISÃO DE TEMPO SEVERO CONVECTIVO UTILIZANDO INTELIGÊNCIA COMPUTACIONAL

JIMMY NOGUEIRA DE CASTRO

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM METEOROLOGIA DO INSTITUTO DE GEOCIÊNCIAS DO CENTRO DE CIÊNCIAS MATEMÁTICAS E DA NATUREZA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO (PPGM-IGEO-CCMN-UFRJ) COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA À OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS (ÁREA: METEOROLOGIA).

PROF. DR. GUTEMBERG BORGES FRANÇA (ORIENTADOR)

PROF. DR. MANOEL VALDONEL DE ALMEIDA

DR. VINÍCIUS ALBUQUERQUE DE ALMEIDA

PROF. DR. HAROLDO FRAGA DE CAMPOS VELHO

PROF. DR. WALLACE FIGUEIREDO MENEZES

RIO DE JANEIRO, RJ - BRASIL

OUTUBRO DE 2021

Programa de Pós-Graduação em Meteorologia (PPGM), Instituto de Geociências (IGEO),
Centro de Ciências Matemáticas e da Natureza (CCMN), Universidade Federal do Rio de Janeiro (UFRJ).
Av. Athos da Silveira Ramos, 274, Bloco G, Cidade Universitária, Ilha do Fundão, Rio de Janeiro, RJ, Brasil, CEP:
21941-916.

Tel.: 55 (21) 3938-9540. E-mail: "Programa de Pós-graduação em Meteorologia - PPGM"
<pos_meteoro@igeo.ufrj.br>

AGRADECIMENTOS

Em primeiro lugar, agradeço a Deus e aos meus pais, Carlos Alberto e Vilma, por me guiarem até essa longa caminhada que, com incansável dedicação e apoio, sempre me forçando a andar para frente e nunca recuar.

Aos meus filhos, Pedro, Gabriela e Vinícius, minha maior motivação e amor que norteiam a minha vida.

A minha esposa, Cristina, pelo amor, paciência e dedicação nos momentos de ausência e devoção.

Aos meus irmãos, Julian e Jeimis, pelo exemplo, eterna amizade e apoio constate nas dificuldades e desafios superados e a superar.

Ao meu amigo Fabrício, meu companheiro de mestrado, pelo apoio constante e pela força nos momentos difíceis.

Ao Departamento de Controle do Espaço Aéreo (DECEA), por ter proporcionado a minha participação neste curso de Pós-Graduação.

A UFRJ, por me acolher e prover os recursos apropriados para o curso de excelência que tive o privilégio de participar.

A meus amigos, Carol e Alexandre, pela disponibilidade, incentivo, apoio e pelos divertimentos que tornaram a nossa caminhada menos sacrificante.

Aos Professores da UFRJ, ao Dr. Gilberto F. Fisch, ao Dr. Haroldo F. de Campos Velho e Dr. Manoel Valdonel de Almeida, pelos ensinamentos e dedicação que permitiram a abertura de janelas do saber, antes, inexistentes para a minha visão.

E, em especial, ao Prof. Gutemberg B. França, pela extrema paciência, dedicação, acolhimento, doação e, principalmente, pelos ensinamentos que ultrapassam a esfera acadêmica.

RESUMO

Este trabalho desenvolveu modelos, baseado em aprendizado de máquina, de previsão de 6 horas de ocorrência de tempo severo convectivo na região de aproximação de pouso dos aeroportos nas circunvizinhanças de São Paulo. No treinamento e teste dos algoritmos de aprendizado de máquina foram utilizados os índices termodinâmicos clássicos (*input*), obtidos dos perfis atmosféricos das radiossondagens de Marte-São Paulo e descargas atmosférica para caracterização do tempo severo convectivos (*output*), considerando período de 2001 a 2017. A análise do histórico dos eventos definiu as áreas e horários de maiores ocorrências e a severidade dos eventos convectivos. As estatísticas POD, 1-FAR, BIAS, KAPPA e F-MEASURE foram utilizadas para avaliar a previsão de 6 horas da detecção de evento convectivo (e entre parênteses para se é severo, quando há ocorrência de descargas atmosférica superior ou igual 1.000), apresentando, respectivamente, valores iguais a 0,91(0,85), 0,95(0,94), 0,92(0,89), 0,74(0,77), e 0,88(0,95). Os resultados da aplicação do modelo em trinta dias aleatórios (*handcast*) obtiveram 96,7% de acertos (excetos dois eventos) dos eventos convectivos e 86,7 se eles são severos. Dois erros de detecção do evento do modelo são discutidos e indicaram que os eventos não foram severos e suas gêneses associados a entrada de sistemas frontais.

Palavras-chave: eventos meteorológicos convectivos, inteligência computacional, índices de instabilidade atmosféricos, descargas atmosféricas.

SEVERE CONVECTIVE WEATHER FORECAST USING COMPUTATIONAL INTELLIGENCE

ABSTRACT

This work developed models, based on machine learning, for the 6-hour forecast of severe convective weather in the landing approach region of airports in the vicinity of São Paulo. In the training and testing of the machine learning algorithms, the classical thermodynamic indices (*input*) were used, obtained from the atmospheric profiles of the Marte-São Paulo radiosonde and atmospheric discharges to characterize convective severe weather (*output*), considering the period from 2001 to 2017. The analysis of the history of events defined the areas and times of greatest occurrences and the severity of convective events. The POD, 1-FAR, BIAS, KAPPA and F-MEASURE statistics were used to evaluate the 6-hour forecast of convective event detection (and in parentheses for whether it is severe, when there is an occurrence of lightning greater than or equal to 1,000), presenting, respectively, values equal to 0.91(0.85), 0.95(0.94), 0.92(0.89), 0.74(0.77), and 0.88(0.95). The results of the application of the model in thirty random days (handcast) obtained 96.7% of correct answers (except for two events) of convective events and 86.7 if they are severe. Two model event detection errors are discussed and indicated that the events were not severe and their genesis associated with frontal systems input.

keywords: convective meteorological events, machine learning, atmospheric instability indexes, lightning strikes.

SUMÁRIO

1. INTRODUÇÃO	12
1.1 MOTIVAÇÃO	13
1.2 OBJETIVO	15
2. REVISÃO BIBLIOGRÁFICA.....	16
3. FUNDAMENTAÇÃO TEÓRICA	19
3.1 TEMPESTADES CONVECTIVAS SEVERAS	19
3.2 DESCARGAS ATMOSFÉRICAS	20
3.3 ÍNDICES DE INSTABILIDADE TERMODINÂMICOS (IIT).....	23
3.3.1 SHOWALTER (SH).....	23
3.3.2 LEVANTAMENTO (LI)	24
3.3.3 TOTAL TOTALS (TT).....	25
3.3.4 SWEAT (SW)	26
3.3.5 ÍNDICE K	26
3.3.6 ÁGUA PRECIPITÁVEL (AP)	27
3.5 INTELIGÊNCIA ARTIFICIAL (IA)	28
3.5.1 SOFTWARE WEKA	29
3.6 ÁREA TERMINAL DE AERÓDROMO (TMA).....	29
4. DADOS E MÉTODO	31
4.1 ÁREAS DE ESTUDO	31
4.2 DADOS	32
4.2.1 DADOS DE DESCARGAS ATMOSFÉRICAS	32
4.2.2 ÍNDICES DE INSTABILIDADE TERMODINÂMICOS	33
4.3 MÉTODO	33
4.3.1 DETALHES DO MÉTODO	34
4.3.3 ESTATÍSTICAS DE AVALIAÇÃO	35
5. RESULTADOS.....	38
5.1 ANÁLISE DOS DADOS.....	38
5.1.1 DADOS DE RADIOSSONDAGENS	38

5.1.2 DESCARGAS ATMOSFÉRICAS	39
5.3 PREVISÃO DE EVENTO CONVECTIVO.....	42
5.4 CLASSIFICAÇÃO DA SEVERIDADE DO EVENTO CONVECTIVO	43
5.4 APLICAÇÃO: ESTUDO DE CASOS	46
5.4.1 ANÁLISE DO DIA 22 DE FEVEREIRO DE 2013	48
5.4.2 ANÁLISE DO DIA 13 DE MARÇO DE 2013.....	48
6. CONCLUSÃO	50
8. REFERÊNCIAS.....	51
ANEXO I	56
ANEXO II	60

LISTA DE FIGURAS

Figura 1.1: Principais fenômenos meteorológicos que impactam a aviação durante as fases do voo.	12
Figura 1.2: Crescimento do Tráfego Aéreo Global.	13
Figura 1.3: Trajetória dos Voos TAM 3096 com a imagem do Satélite GOES-13 Realçada das 1800Z (a) e do G31113 do dia 27/01/2016 com a imagem do Radar do dia 27/01/2016 das 1800Z (b).	14
Fig. 3.1: TMA-SP.	30
Figura 4.1: Área de estudo.	32
Figura 4.2: Fluxograma do método utilizado na pesquisa.	33
Figura 4.3: Total de radiossondas por ano disponíveis para esta pesquisa.	39
Figura 5.1: Densidade de ocorrências de DA no período de 2000 a 2017.	41
Figura 5.2: Comportamento dos índices atmosféricos SH (a), LI (b), SW (c), K (d), TT (e) e AP (f) versus o intervalo do total de ocorrências de DA.	43
Figura 5.3: Distribuição mensal (a) e horária (b) de ocorrências de DA no período de dezembro de 2000 a junho de 2017.	44
Figura 5.4: Imagens de Satélite GOES-13, do canal Infravermelho (CH04), do dia 22 de fevereiro de 2013 das 1230Z (a) e 2000Z (b). Fonte: Sítio de Internet do CPTEC/INPE, acessado em 06 de maio de 2021.	48
Figura 5.5: Imagens de Satélite GOES-13, do canal Infravermelho (CH04), do dia 22 de fevereiro de 2013 das 1230Z (a) e 1800Z (b). Fonte: Sítio de Internet do CPTEC/INPE, acessado em 06 de maio de 2021.	49

LISTA DE TABELAS

Tabela 3.1: Valores de SH ajustados para latitudes médias. Fonte	24
Tabela 3.2: Valores do índice de levantamento (LI).....	25
Tabela 3.3: Valores de TT para tempestades.	25
Tabela 3.4: Valores de SW para latitudes médias.....	26
Tabela 3.5: Valores de K para tempestades.	27
Tabela 4.1: Dados utilizados pela pesquisa.	32
Tabela 4.2: Matriz de confusão.	35
Tabela 4.3: Interpretação para os valores do índice KAPPA.	36
Tabela 5.1: Ocorrência mensal de da na área da pesquisa do período de 2000 a 2017.	40
Tabela 5.2: Valores das estatísticas dos melhores resultados iniciais da previsão-5h(15-00h) de evento convectivo (ou não, e sua severidade (quantidade de DA).	42
Tabela 5.3: Resultado dos modelos preditos.	46
Tabela 5.4: Resultado do estudo de caso com 30 amostras aleatórias.	47

LISTA DE SIGLAS E ABREVIATURAS

AM - Aprendizado de Máquina

AP - Índice atmosférico termodinâmico denominado Água Precipitável

BIAS - Proporção da Frequência dos Eventos Preditivos pela Frequência dos Eventos Observados

CEMIG - Companhia Energética de Minas Gerais

CGNA - Centro de Gerenciamento Nacional da Navegação Aérea

COPEL - Companhia Paranaense de Energia

DA - Descarga Atmosférica

FAR - Probabilidade de Falsa Detecção ou alarme falso

F-MEASURE - Medida Harmônica entre a Precisão e a Sensibilidade

IA - Inteligência Artificial

IIT - Índices de Instabilidade Termodinâmicos

INPE - Instituto Nacional de Pesquisas Espaciais

INPUT - Entrada

K - Índice atmosférico termodinâmico K

KAPPA - Medida de Concordância entre o Previsto e o Observado

LI - Índice atmosférico termodinâmico LIFT

LLS - Lightning Locating Systems

NM – Milha Náutica

OACI - Organização de Aviação Civil Internacional

OUTPUT - Saída

POD - Probabilidade de Detecção do Evento

RAMS - Regional Atmospheric Modeling System

RINDAT - Rede Integrada Nacional de Detecção de Descargas Atmosféricas

SBGR - Indicador de localidade da OACI para o Aeroporto de Guarulhos

SBKP - Indicador de localidade da OACI para o Aeroporto de Campinas

SBMT - Indicador de localidade da OACI para o Aeroporto de Marte

SBSJ - Indicador de localidade da OACI para o Aeroporto de São José dos Campos

SBSP - Indicador de localidade da OACI para o Aeroporto de Congonhas

SBST - Indicador de localidade da OACI para o Aeroporto de Santos

SH – Índice atmosférico termodinâmico SHOWALTER

SIMEPAR - Sistema Meteorológico do Paraná

SISCEAB - Sistema de Controle do Espaço Aéreo

SW - Índice atmosférico termodinâmico SWEAT

TMA - Área Terminal de Aeródromo

TSC - Tempo Severo Convectivo

TT - Índice atmosférico termodinâmico TOTAL TOTALS

UTC - Tempo Universal Coordenado

WEKA - Waikato Environment for Knowledge Analysis

1. INTRODUÇÃO

Com o contínuo crescimento da aviação global, conhecer com antecedência, as possíveis adversidades nas diversas fases do voo, torna-se o maior desafio para a garantia da segurança e a eficiência das operações aéreas. Dentre as principais adversidades, os fenômenos meteorológicos, quando severos, são os que mais impactam a aviação.

Dentre os principais fenômenos meteorológicos, os associados ao tempo severo convectivo (TSC), ver Figura 1.1, que pode estar associado a chuva forte, e/ou granizo, turbulência, e/ou vento de rajada e forte, e/ou visibilidade baixa e trovoada, destacam-se por ocorrerem em todas as fases de um voo, tornando-se o principal alvo de estudo para o aprimoramento de suas previsões.

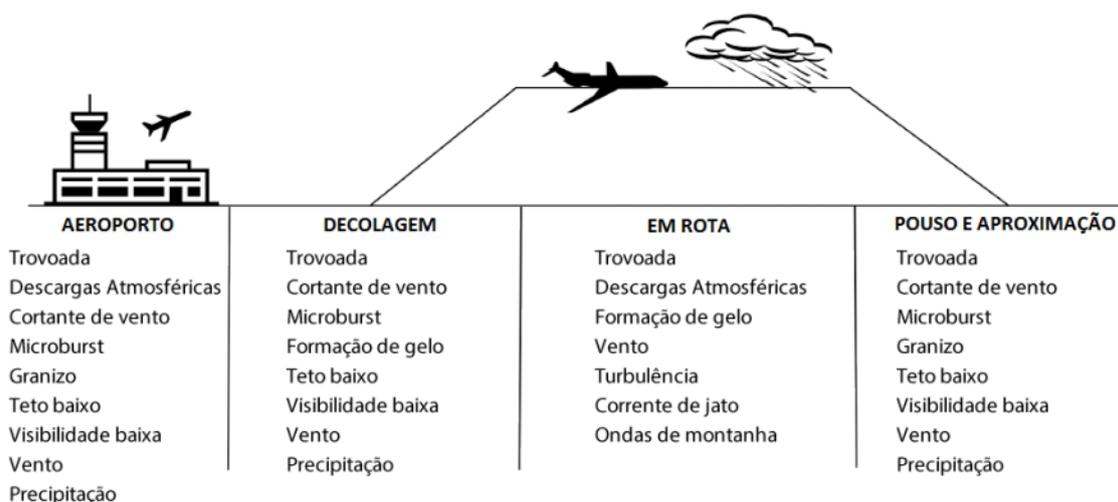


Figura 1.1: Principais fenômenos meteorológicos que impactam a aviação por fases do voo.

Fonte: adaptado de PEJOVIC (2008).

A ocorrência de tempestades severas, oriundas de nuvens convectivas, nas áreas do espaço aéreo destinada aos procedimentos das aeronaves para decolagens e aproximação pousos, denominada *Terminal Manoeuvring Area* (TMA), geram, algumas vezes, expressivos impactos ao gerenciamento do tráfego aéreo, contribuindo, frequentemente, com desvios de rotas, ocasionando atrasos e cancelamentos de voos, aumento do consumo de

combustível e da emissão de CO₂ na atmosfera, que é diretamente responsável pelo aquecimento global, fato frequentemente observado pelos os Órgãos de Controle do Espaço Aéreo (OACI, 2016).

O aprimoramento das atuais técnicas para a previsão das ocorrências de tempestades severas convectivas, nas TMA, contribuirá, sobremaneira, para o aumento da segurança e da economia dos voos, bem como da eficiência da aviação.

1.1 MOTIVAÇÃO

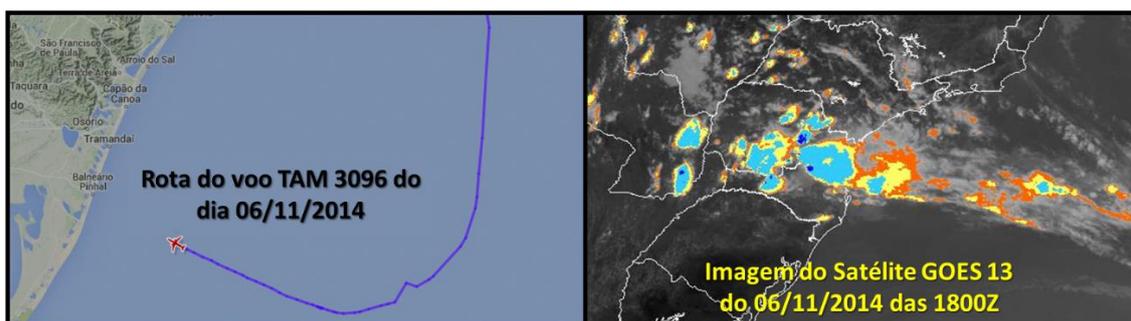
De acordo com a Organização de Aviação Civil Internacional (OACI) e seu Plano de Navegação Aérea Global (OACI, 2016), que visa estabelecer o planejamento para a adaptação da aviação global ao crescimento do tráfego aéreo desde 1977 até 2030, o tráfego aéreo global vem duplicando a cada 15 anos, conforme apresentado na Figura 1.2. Este crescimento, levou a OACI a criar, por intermédio de um Sistema de Melhoria por Blocos, contido nesse plano, diversas metas de implementação para a adaptação da aviação global, visando o contínuo aumento da demanda do espaço aéreo.

Das principais metas estabelecidas pela OACI para a adaptação à aviação do futuro, a meteorologia aeronáutica se destaca com diversas implementações visando o apoio à tomada de decisão durante o voo e no seu planejamento, tendo como objetivo a otimização das operações aéreas, como por exemplo a diminuição do tempo de voo com rotas mais curtas e a diminuição da emissão de CO₂ na atmosfera.

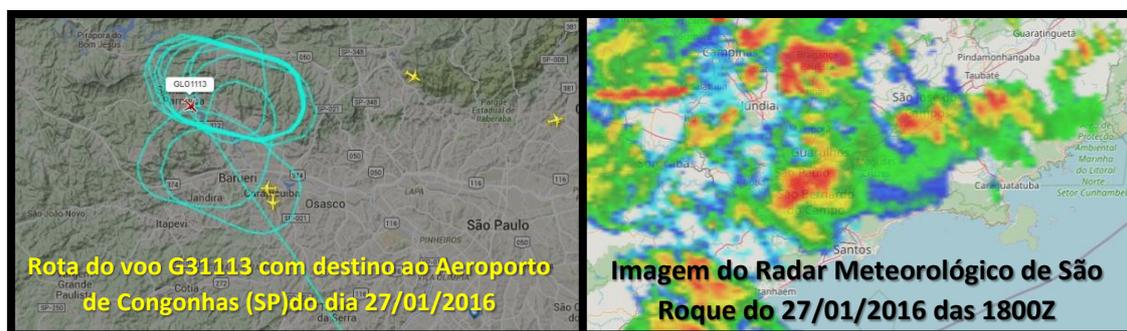


Figura 1.2: Crescimento do Tráfego Aéreo Global. Fonte: OACI (2016)

Na rotina do Centro de Gerenciamento Nacional da Navegação Aérea (CGNA), órgão pertencente ao Sistema de Controle do Espaço Aéreo (SISCEAB) responsável pelo gerenciamento do fluxo de tráfego aéreo brasileiro, é comum a observação de diversos casos de desvios de rotas por ocorrências de tempo severo convectivo, conforme exemplos apresentados na Figura 1.3 (a-b).



(a)



(b)

Figura 1.3 (a-b): Trajetória dos Voos TAM 3096 (contornando TSC) com a imagem do Satélite GOES-13 Realçada das 1800Z do dia 06 de novembro de 2014 (a) e do voo G31113 do dia 27 de janeiro de 2016 (orbitando com objetivo de pousar no Aeroporto de Guarulhos) com a imagem do Radar das 1800Z (b). Fonte: Adaptado de <https://www.flightradar24.com/51,-2/6> e <https://www.redemet.aer.mil.br/stsc/public/produto> e http://satellite.cptec.inpe.br/repositoriogoes/goes13/goes13_web/ams_real_cada_baixa/2014/01/S11232957_201401061800.jpg.

Dos dez aeródromos brasileiros com o maior fluxo de aeronaves, quatro pertencem à TMA de São Paulo e correspondem por cerca de 52% do total de

voos contabilizado (DECEA, 2019). Por esta área demonstrar ser a de maior importância para a aviação brasileira, esta tornou-se foco desta pesquisa.

1.2 OBJETIVO

Diante do impacto causado pelas ocorrências de TSC nos procedimentos de aproximação, pouso e decolagem das aeronaves e na busca do aprimoramento dos prognósticos de TSC, além da TMA-SP ser a área de maior fluxo de tráfego aéreo na aviação brasileira, esta pesquisa propõe avaliar algoritmos de aprendizado de máquina (AM) classificatório para o desenvolvimento de modelo¹ preditivo de eventos convectivo severos na TMA-SP, considerando a densidade de descargas atmosféricas (DA) e os índices de instabilidade termodinâmicos (IIT) extraídos dos perfis atmosféricos.

¹ Modelo representa algoritmo de aprendizado de máquina, ou inteligência computacional, treinado e testado para o propósito definido.

2. REVISÃO BIBLIOGRÁFICA

As tempestades convectivas vêm sendo estudadas há algumas décadas. A maioria desses estudos utilizaram dados de radares meteorológicos como principal fonte para observação das células convectivas associadas às tempestades severas, como trabalhos publicados por Battan (1959), Newton e Fankhauser (1964) e Wilson (1966). Por outro lado, por intermédio de modelos determinísticos, Weisman e Klemp (1986) utilizaram os índices termodinâmicos atmosféricos determinados com o cisalhamento do vento usando dados de radiossondagens para diferenciar os tipos de tempestades. Posteriormente, diversos estudos utilizaram o Sistema de Detecção de DA (LLS - Lightning Locating Systems) para o monitoramento das tempestades convectivas severas, como Turman e Tettelbach (1980), Williams *et al.* (1989) e Kane (1991).

Mccann (1992) propôs um método de previsão de três a sete horas de ocorrência de tempestades severas utilizando duas redes neurais a partir de variáveis baseadas na convergência de umidade da superfície, o que resultou na análise sobre os padrões que antecedem os eventos severos, além de demonstrar o potencial impacto das redes neurais em previsões de eventos meteorológicos.

Mueller *et al.* (2003) desenvolveram um sistema de previsão de até uma hora, com objetivo de localizar as áreas de ocorrência de tempestades, por intermédio de dados de superfície, radar, dados de satélite e modelagem numérica, utilizando os estágios de maturidade das tempestades.

Isaac *et al.* (2006, 2011, 2014) apresentaram diversos trabalhos relacionados à aviação para a previsão de até seis horas para os principais aeroportos do Canadá, para ocorrências de tempo restritivos à aviação, incluindo tempestades severas, com base nos dados de modelos numéricos, observações de superfície, radar, satélite e um radiômetro de micro-ondas.

Nascimento (2005) apresentou um trabalho sobre a previsão de tempo severo no Brasil, destacando parâmetros atmosféricos aplicáveis a identificação da ocorrência de eventos de tempestades convectivas severas, por intermédio das saídas dos modelos de mesoescala. Ainda, Nascimento

(2005) demonstrou que alguns parâmetros utilizados nas latitudes médias do hemisfério norte para a previsão de tempo severo podem ser aplicáveis nas previsões de sistemas convectivos no Brasil, permitindo, também, a utilização destes parâmetros para a elaboração de índices melhorados para o hemisfério sul. Nascimento (2005) afirmou que a utilização dos registros de DA contribui, fortemente, para as previsões de *nowcasting* e para o monitoramento do tempo.

No Brasil, há vários estudos sobre a previsão de tempestades severas. Queiroz (2009) desenvolveu o índice de severidade para a previsão de curto prazo (*nowcasting*) de tempestades convectivas severas utilizando dados de radar meteorológico e a ferramenta FORTRACC.

Grossman (2010) determinou, em seu trabalho, os locais de maiores incidências de ocorrências de tempestades, por intermédio de simulações numéricas no estado do Rio de Janeiro.

França et al. (2016), utilizando redes neurais, apresentaram um modelo de previsão de locais de instabilidade atmosférica significativa para TMA-RJ, utilizando dados de observações meteorológicas de superfície, perfis atmosféricos (gerados pelas radiossondagens), dados pluviométricos e registros de DA.

Paulucci (2017), analisando um período de dezesseis anos (2001-2016) das ocorrências de eventos de DA do tipo nuvem-solo, na área metropolitana do Rio de Janeiro, observando a polaridade, correntes de pico, distribuições geográficas e variabilidade diurna, interanual e interanual, identificou que as ocorrências de DA, na área estudada, são concentradas nos períodos do verão (64,3%) e primavera (20,5%) e, nos períodos do outono (12,9%) e inverno (2,3%) de menor incidência, além de identificar os horários de 18h e 19h como os de maiores ocorrências e que a maioria dos raios observados derivou do efeito orográfico, que estimula a formação de tempestades convectivas ao longo da parte sul da encosta.

Freitas et al. (2018) utilizaram IIT, com análises de ocorrências de DA, em uma árvore de decisão para previsão de tempestades severas na Região Metropolitana do Rio de Janeiro, obtendo resultados satisfatórios que

indicavam taxa de acertos geral acima de 80% e taxa de erro geral abaixo de 20% na previsão dos eventos.

Ainda, Almeida *et al.* (2020), desenvolveram um método, baseado em algoritmo de AM, para a previsão de eventos convectivos meteorológicos na TMA do Galeão, na cidade do Rio de Janeiro. Os resultados revelaram o potencial da inovadora modelagem empregada no *nowcasting* dos eventos estudados, uma vez que as estatísticas probabilidade de detecção, falso alarme e viés possuem valores, respectivamente, valores iguais a 0,91, 0,85 e 1,03.

3. FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta a revisão da literatura, abordando os assuntos desta dissertação: como tempestades convectivas severas, descargas atmosféricas, índices de instabilidade termodinâmicos, Inteligência Artificial e a Área Terminal de Aeródromo, conforme descritos nas seções a seguir.

3.1 TEMPESTADES CONVECTIVAS SEVERAS

A tempestade convectiva severa é um fenômeno meteorológico associado, em sua origem, à maturação da nuvem *cumulonimbus*, sendo capaz de gerar DA, tornados, rajadas de ventos fortes com efeitos devastadores e precipitações de granizo e chuva forte.

Segundo Riehl (1965), o ciclo evolutivo de uma nuvem *cumulonimbus* é classificado por três estágios, conforme descritos abaixo:

- Estágio Cúmulos: Fase inicial, caracterizada como o início da formação da nuvem com a ascensão do ar;
- Estágio Maduro: Esta fase é onde ocorre os fenômenos severos associados, como a precipitação de chuvas fortes e granizo, além de ventos fortes; e
- Estágio de Dissipação: Este é o último estágio da nuvem. Seu início caracteriza-se com as correntes descendentes frias atingindo o solo, a chuva resfria o ar nos níveis mais baixos e cessam-se novas fontes de instabilidade.

De acordo com Orlanski (1975), tempestades severas são sistemas convectivos de mesoescala capazes de gerar granizo de tamanho grande (com pedras de 2 cm ou mais de diâmetro ao atingirem a superfície), vendavais intensos com força destrutiva (velocidade acima de 26 metros por segundo) ou tornados.

Varejão-Silva (2006) definiu a nuvem *cumulonimbus* como volumosa e densa, com grande desenvolvimento vertical, podendo apresentar uma imensa bigorna ou penacho em sua parte superior, com extremidades cirrosas e com

potencial de produzir aguaceiros violentos, acompanhados de relâmpagos e trovões, além de granizo ou saraiva.

3.2 DESCARGAS ATMOSFÉRICAS

As DA, comumente denominadas de relâmpagos, de maneira simples, podem ser definidas como disparos elétricos de grande intensidade com duração de meio segundo e trajetória com comprimento de 5 a 10 Km, aproximadamente e podendo ultrapassar um quilo ampère. Sua formação se dá pela fricção das partículas dentro de uma nuvem *cumulonimbus* ou entre duas nuvens *cumulonimbus* próximas, tornando-a carregada eletricamente. Ela pode ocorrer dentro de uma nuvem, entre duas nuvens ou entre uma nuvem e o solo (ELAT, 2021).

Os estudos iniciais de DA se deram a partir da descoberta da eletricidade no início do século XVIII. Em 1889, H.H. Hoffert registrou, como uma câmera fotográfica, um relâmpago próximo ao solo. Em 1897, F. Pockels, pioneiramente, estudou sobre a intensidade máxima de corrente de um relâmpago próximo ao solo, por medidas de campo magnético residual da ocorrência de DA em rochas basálticas (ELAT, 2021).

Nos dias atuais, com o aprimoramento dos estudos sobre a eletricidade atmosférica, foi possível avançar no entendimento da estrutura elétrica da atmosfera. Conforme ELAT (2021), esta estrutura se dá pela mistura de gases e partículas (como o nitrogênio, oxigênio e vapor d'água), pela relação dos íons e a condutividade e pela temperatura, conforme Figura 3.1.

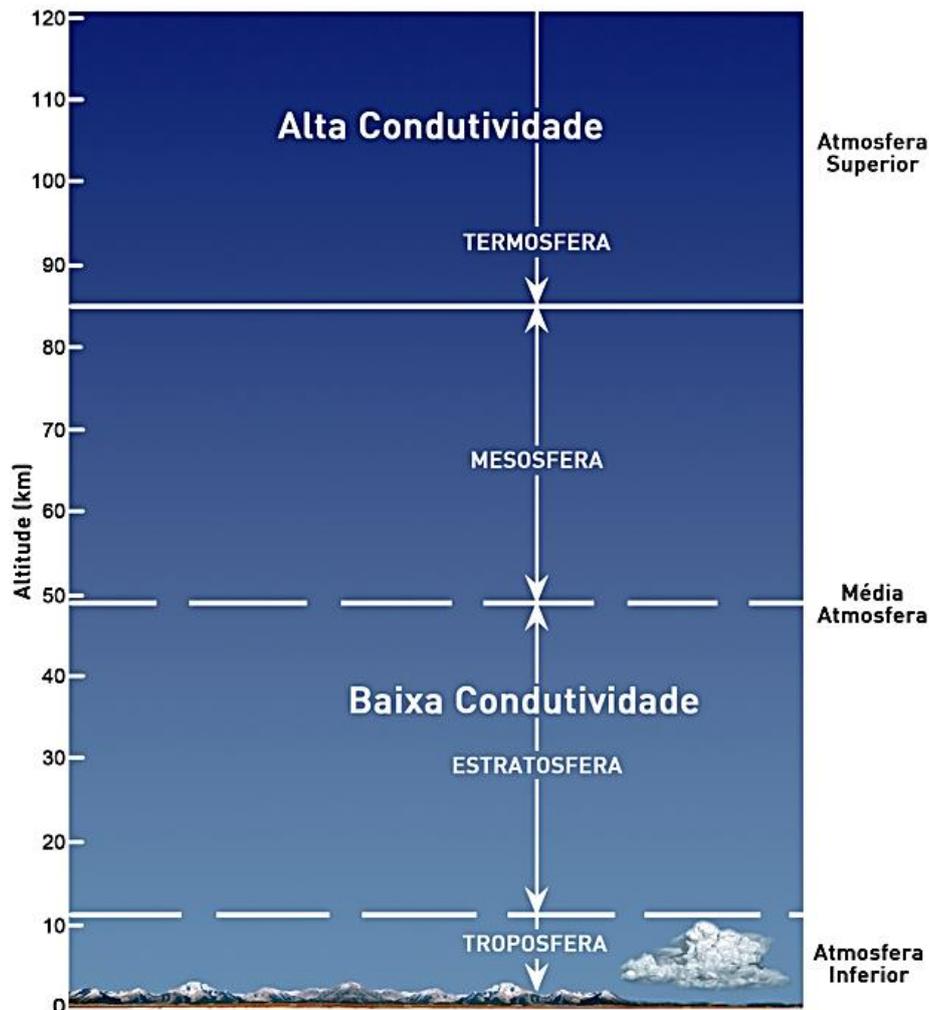


Figura 3.1: Estrutura da atmosfera quanto a sua condutividade e temperatura. Fonte: sítio de Internet do ELAT <<http://www.inpe.br/webelat/homepage/menu/relamp/eletricidade.atmosferica/estrutura.eletrica.da.atmosfera.php>>, acessado em 06 de maio de 2021.

Conforme o ELAT (2021), de forma geral, os relâmpagos podem ser classificados de duas formas, a saber: relâmpagos no solo (mais comuns) e relâmpagos na nuvem (raros). Quanto ao sinal de carga líquida transportada da nuvem para o solo, os relâmpagos podem ser classificados como: negativos ou positivos, conforme Figura 3.2.

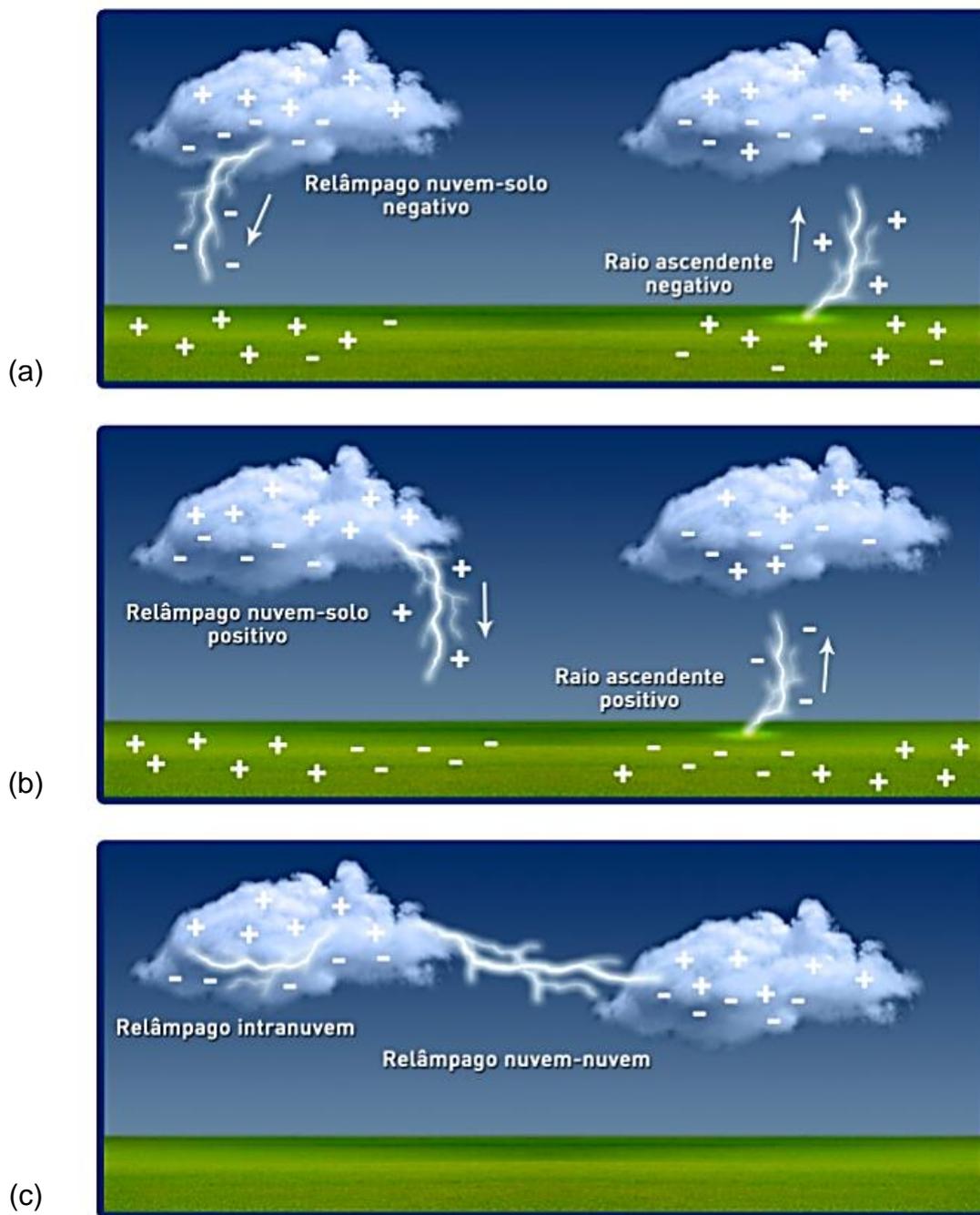


Figura 3.2: Representação dos relâmpagos nuvem-solo (a-b) e intranuvem (c). Fonte: sítio de Internet do ELAT <<http://www.inpe.br/webelat/homepage/menu/relamp/eletricidade.atmosferica/estrutura.eletrica.da.atmosfera.php>>, acessado em 06 de maio de 2021.

Para o estudo do prognóstico de TSC, os relâmpagos, aqui denominados de DA, funcionam como parâmetro localizador e caracterização da severidade das nuvens cumulonimbus e que podem ser monitorados por intermédio dos

Sistemas de Localização de DA, como é o caso do Sistema RINDAT (Rede Integrada Nacional de Descargas Atmosféricas, <http://www.inpe.br/webelat/rindat/>).

3.3 ÍNDICES DE INSTABILIDADE TERMODINÂMICOS (IIT)

A instabilidade atmosférica é fator essencial para ocorrência de TSC. O grau dessa instabilidade pode ser identificado por meio de diagramas termodinâmicos, analisando-se as curvas de estado da atmosfera em comparação à evolução de uma parcela de ar, levando-se em conta os gradientes adiabáticos. No entanto, os IIT, valores determinados a partir dos perfis verticais de temperatura e umidade da atmosfera registrados usando de sondagens atmosféricas (radiossondagens), foram idealizados para traduzir numericamente o grau de instabilidade atmosférica. Em outras palavras, os IIT auxiliam na determinação do potencial convectivo da atmosfera, sendo utilizados, portanto, como uma das principais ferramentas para os prognósticos do TSC, a exemplo do estudo desenvolvido por Mourão e Menezes (2006), que utilizaram com sucesso simulações do Modelo Regional RAMS (*Regional Atmospheric Modeling System*) e alguns IIT para previsão de tempestades convectivas no Estado do Rio de Janeiro.

Nesta pesquisa, seis IIT, descritos sucintamente abaixo, são utilizados como variáveis preditoras dos algoritmos de AM a serem treinados e avaliados, isto é: SHOWALTER (SH), LIFT (LI), TOTAL TOTALS (TT), SWEAT(SW), K e Água Precipitável (AP).

3.3.1 SHOWALTER (SH)

O SH é utilizado para medir o potencial e a intensidade de uma tempestade. O SH foi um dos primeiros índices desenvolvidos, utilizando a temperatura do ar e do ponto de orvalho em 850 hPa e da temperatura do ar em 500 hPa (SHOWALTER, 1947).

Quanto mais negativo forem os valores deste índice, maior será o potencial da atmosfera para ocorrência de tempestades.

O SH é calculado a partir da Eq. 3.1,

$$SH = T_{500} - T_{parcela}; \quad (\text{Eq. 3.1})$$

onde, T_{500} representa a temperatura do ar (em graus Celsius) em 500 hPa e $T_{parcela}$ representa a temperatura em 500 hPa (em Celsius) de uma parcela que subiu desde o nível de 850 hPa

Os valores utilizados para a classificação da possibilidade de ocorrências da formação de tempestades são descritos na Tabela 3.1.

Tabela 3.1: Valores de SH ajustados para latitudes médias. Fonte: (SHOWALTER, 1947).

SH	Possibilidade de Tempestades
IS > +3	Sem Atividade Convectiva
+1 ≤ IS ≤ +3	Possíveis Pancadas de Chuva/Tempestades Isoladas
-2 ≤ IS ≤ +1	Tempestades Prováveis
-6 ≤ IS ≤ -2	Possibilidade de Tempestades Severas
IS ≤ -6	Tempestades Severas Prováveis/Possibilidade de Tornados

3.3.2 LEVANTAMENTO (LI)

O LI, do inglês *LIFT*, é expresso pela diferença da temperatura ambiente e da temperatura de uma parcela de ar levantada adiabaticamente a uma determinada altura de pressão na troposfera, geralmente 500 hPa. Quando o valor é positivo, a atmosfera (na altura respectiva) é estável e, quando o valor é negativo, a atmosfera é instável e quanto menor for este índice, maior será o potencial para ocorrência de tempestade.

O LI é representado na Eq. 3.2:

$$LI = T_{500} - T_{p500} [^{\circ}\text{C}]; \quad (3.2)$$

onde, T é a temperatura do ambiente e T_p é a temperatura da parcela em 500 hPa. No LI a parcela é levantada da superfície. Este é definido na Tabela 3.2.

Tabela 3.2: Valores do Índice de Levantamento (LI) (GALWAY, 1956).

LI	Nível de Instabilidade
$0 \leq IL \leq +3$	Estável. Possível convecção fraca na presença de forte levantamento ou algum mecanismo forçante
$-3 \leq IL \leq 0$	Marginalmente Instável
$-6 \leq IL \leq -3$	Marginalmente Instável
$-9 \leq IL \leq -6$	Muito Instável
$IL < -9$	Extremamente Instável

3.3.3 TOTAL TOTALS (TT)

O TT é utilizado como ferramenta de análise do entranhamento de ar frio na troposfera média, sendo importante para previsão de tempestades severas com ocorrência de precipitação de granizo.

O TT pode ser calculado a partir da Eq.3.3:

$$TT = T_{d850} + T_{850} - 2(T_{500}); \quad (\text{Eq. 3.3})$$

onde T é a temperatura do bulbo seco e T_d a temperatura do ponto de orvalho (MILLER, 1972).

A representação dos valores do índice TT é apresentada conforme a Tabela 3.3.

Tabela 3.3: Valores de TT para Tempestades (MILLER, 1972).

TT	Potencial para Tempestades
$40 \leq TT \leq 45$	Isoladas, algumas poucas moderadas
$46 \leq TT \leq 47$	Esparsas, algumas poucas pesadas
$48 \leq TT \leq 49$	Esparsas moderadas, algumas pesadas e algumas isoladas severas
$50 \leq TT \leq 51$	Esparsas pesadas, algumas severas e tornados isolados
$52 \leq TT \leq 55$	Esparsas a numerosas pesadas, poucas a esparsas pesadas, alguns tornados
$TT > 55$	Esparsas a numerosas pesadas, poucas a esparsas pesadas, alguns tornados

3.3.4 SWEAT (SW)

Este utiliza informação sobre as estruturas termodinâmica e cinemática da atmosfera. Sua aplicação é indicada para o prognóstico do potencial de tempestades severas. Valores do SW acima de 300 são considerados indicativos de situações propícias a tempestades severas, enquanto os valores acima de 400 indicam maior probabilidade de eventos de tornados (MILLER, 1972).

O SW é calculado a partir da umidade nos baixos níveis, pelo ponto de orvalho em 850 hPa, pela instabilidade, através do índice TT, pelo jato de baixos níveis, obtido do vento em 850 hPa, pelo jato de altos níveis, obtidos pelo vento em 500 hPa e pela advecção quente, entre 850 e 500 hPa. A Eq. 3.4 apresenta matematicamente o SW.

$$SW = 12Td_{850} + 20(TT-49) + 2V_{850} + V_{500} + 125f(C); \quad (\text{Eq. 3.4})$$

onde, $C = \text{sen}(dir_{500} - dir_{850}) + 0,2$, dir_{500} representa a direção do vento em 500 hPa e dir_{850} representa a direção do vento em 850 hPa.

Tabela 3.4: Valores de SW para latitudes médias (MILLER, 1972).

SW	Potencial para Tempestades Severas
$SW \leq 300$	Sem potencial para tempestades severas
$300 \leq SWEAT \leq 400$	Potencial para tempestades severas
$SWEAT > 400$	Potencial para tornados

3.3.5 ÍNDICE K

Este expressa o potencial para a ocorrência de tempestades e fundamentando-se na taxa de variação vertical da temperatura, no conteúdo de umidade na baixa troposfera e na extensão da camada úmida.

O índice K é definido como na Eq. 3.5 (GEORGE, 1960):

$$K = (T_{850} - T_{500} + T_{d850} - (T_{700} - T_{d700})); \quad (\text{Eq. 3.5})$$

onde T_{850} é temperatura do nível de 850 hPa, Td_{850} expressa a temperatura do ponto de orvalho do nível de 850 hPa, T_{700} representa a temperatura do nível de 700 hPa, Td_{700} indica a temperatura do ponto de orvalho do nível de 700 hPa e T_{500} representa a temperatura do nível de 500 hPa ($^{\circ}\text{C}$);

De acordo com Henry (1987), o índice K com valores acima de 30 são considerados com alto potencial para a ocorrência de tempestades e acima de 40, extremamente altos.

O K é utilizado para prognósticos de trovoadas de massas de ar, ocorrendo com o resultado do movimento vertical de uma parcela de ar, devido ao aquecimento local ou efeito orográfico que essa parcela está sujeita ao elevar-se pela encosta de uma montanha. A Tabela 3.5 mostra os valores de K para tempestades.

Tabela 3.5: Valores de K para Tempestades (GEORGE, 1960).

K	Possibilidade de Tempestades
$K < 20$	Sem Atividade Convectiva
$20 < K < 25$	Tempestades Isoladas
$25 < K < 30$	Tempestades amplamente dispersas
$30 < K < 35$	Tempestades dispersas
$35 < K < 40$	Muitas Tempestades

3.3.6 ÁGUA PRECIPITÁVEL (AP)

O índice AP representa a quantidade de vapor de água atmosférico contido em uma coluna vertical de área de seção transversal unitária, quantificando o volume líquido, supondo a condensação de todo o valor de água mensurado. O índice AP é expresso pela Eq. 3.6, onde p representa o nível de pressão atmosférica e g a aceleração da gravidade,

$$AP = \frac{1}{\rho g} \int_{p_1}^{p_2} x dp; \quad (\text{Eq. 3.6})$$

onde, $x(p)$ é a razão de mistura no nível de pressão (p) e AP representa o vapor de água precipitável, contido em uma camada limitada pelas pressões p_1 e p_2 .

3.5 INTELIGÊNCIA ARTIFICIAL (IA)

Segundo Munakata (1994) existem várias definições distintas para o termo IA principalmente devido a suas infinidades de aplicações. No dicionário *World College Dictionary*, IA é definida como o estudo dos sistemas computacionais que realizam tarefas que os humanos precisam de inteligência para realizá-las.

De acordo com Haykin (2001), IA foi desenvolvida para, a partir de sistemas computacionais, simular a realização de tarefas cognitivas semelhantes às realizadas pelos seres humanos.

A IA começou a ser estudada no início da Segunda Guerra Mundial por Hebert Simon, Allen Newell, Jonh McCarthy e vários outros, que, com objetivos em comum, tinham a intenção de criar um mecanismo capaz de simular a forma do ser humano pensar e tomar decisões.

Com a evolução da IA, foi desenvolvido ferramentas computacionais para aplicação do Aprendizado de Máquinas (AM), chamado em inglês de *Machine learning*, que se constitui do aprendizado de padrões por intermédio de sistemas computacionais.

De acordo com Yao (2013), o AM é um dos principais ramos da IA voltado para o desenvolvimento de modelos computacionais de aprendizado. O AM pode ser aplicado em várias áreas, como: ciências cognitivas, ciência da computação, estatística, complexidade computacional, teoria da informação, teoria do controle, filosofia e biologia. O AM, do ponto de vista computacional, se refere à capacidade de uma máquina melhorar seu desempenho com base em resultados anteriores. Do ponto de vista biológico, o AM é o estudo de como criar computadores que aprenderão com a experiência e modificarão suas atividades com base nesse aprendizado, em oposição aos computadores tradicionais que suas atividades não mudam a menos que os programadores realizem as mudanças explicitamente.

O estudo realizado por Grewal (2014) enfatizou as diferentes definições de IA considerando todas as visões sobre tema e aplicações ao longo do tempo; e, assim, define que IA tem função de desenvolver métodos para a resolução de

problemas, simulando o cérebro humano, por intermédio de algoritmos e sistemas computacionais.

Para a utilização de AM, existem ferramentas que facilitam o reconhecimento de padrões, como, por exemplo, as redes neurais, que são sistemas computacionais, com nós interconectados, que funcionam como os neurônios do cérebro humano.

Para o treinamento dos modelos preditivos, esta pesquisa utilizou os algoritmos de AM classificatórios, que funcionam como ferramenta para o reconhecimento de padrões. Os algoritmos classificatórios são um modelo prático para a inferência indutiva.

3.5.1 SOFTWARE WEKA

Nesta pesquisa, o pacote de software denominado WEKA (*Waikato Environment for Knowledge Analysis*), de autoria Frank et al. (2016) e desenvolvido na Universidade de Waikato, Nova Zelândia é utilizado, visto que este possui um conjunto de algoritmos clássicos de AM do tipo regressivos e classificatórios. Então, considerando que o problema a ser resolvido na presente pesquisa é do tipo classificatório, ou seja, trata-se em identificar a partir do conjunto de valores do IIT o comportamento futuro da atmosfera local, SE será propício SIM ou NÃO ao surgimento de tempestade convectiva severa. Desta forma, os algoritmos classificatórios disponíveis no WEKA - listados no Anexo I - serão aqui utilizados para construção dos modelos de *nowcasting* de TSC, conforme método descrito na seção 4.3.

3.6 ÁREA TERMINAL DE AERÓDROMO (TMA)

Em aviação, TMA é definida como área de controle terminal, por designar uma região do espaço aéreo, normalmente situada na confluência de rotas de tráfego aéreo e nas imediações de um ou mais aeródromos, onde haja um alto volume de tráfego aéreo.

As TMA, que têm limites laterais e verticais variáveis, são áreas críticas para o controle de tráfego, visto que constituem um espaço aéreo com

aeronaves se aproximando para pouso e outras que acabaram de decolar, exigindo enorme atenção, concentração e assertividade por parte do controlador de tráfego aéreo.

A Figura 3.3 apresenta a TMA-SP, na qual concentra o maior fluxo de tráfego aéreo no Brasil (CGNA, 2019), dividida em duas partes: a principal, reunindo os aeroportos de Guarulhos (SBGR), Marte (SBMT), Congonhas (SBSP), Campinas (SBKP), São José dos Campos (SBSJ) e Santos (SBST) e a área secundária contemplando as aerovias que ligam os estados do Rio de Janeiro e São Paulo, conhecidas como tubulão.

A TMA-SP foi escolhida para o foco desta pesquisa por ser a área de maior fluxo de tráfego aéreo no Brasil (DECEA, 2019) e por ser constantemente impactada por TSC ocasionando, em cascata, atrasos e cancelamentos de voos em todo o Brasil, conforme os *registros dos relatórios de situação diária*² do CGNA catalogados pelos DECEA em 2019.



Fig. 3.3: TMA-SP. Fonte: Autor, 2020.

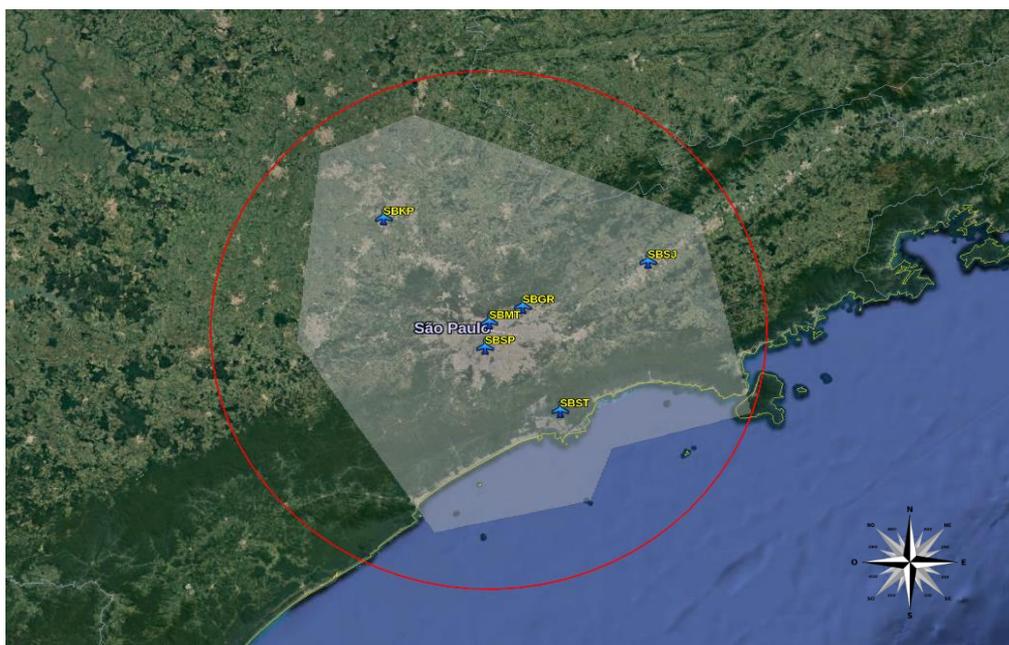
² Refere-se ao livro de registros de ocorrências diárias do gerente nacional (GNAC) do fluxo de tráfego aéreo do CGNA, contendo o relato das medidas de harmonização executadas para a garantia, de forma otimizada, do fluxo de tráfego aéreo nacional.

4. DADOS E MÉTODO

Este capítulo apresenta os detalhes da área de estudo, dos dados utilizados e o método proposto.

4.1 ÁREAS DE ESTUDO

Esta compreende da área principal da TMA-SP, contudo, objetivando delineamento desta área, foi traçado um *círculo de raio de 137 km*, centrado na localização da estação meteorológica de altitude do Campo de Marte com coordenadas geográficas $23^{\circ} 30' 24''$ S e $46^{\circ} 38' 02''$ W, conforme Figura 4.1, compreendendo os aeroportos de Guarulhos (SBGR), Marte (SBMT), Congonhas (SBSP), Campinas (SBKP), São José dos Campos (SBSJ) e Santos (SBST)³.



³ A área de estudo definida encontra-se dentro da área de cobertura das Estações de Altitude, preconizadas pela OMM, conforme *Guide to the GCOS Surface Network (GSN) and GCOS Upper-Air Network (GUAN)*, editado em 2010, disponível em https://library.wmo.int/doc_num.php?expl_num_id=3855.

Figura 4.1: Área de estudo. Fonte: Autor, 2020.

4.2 DADOS

Discorre sobre os dados utilizados os quais são sumarizados na Tabela 4.1.

Tabela 4.1: Dados utilizados pela pesquisa.

TIPO DE DADOS	FINALIDADE	FREQUÊNCIA	PERÍODO	QUANTIDADE	FONTE
23 índices conforme Anexo II	Variáveis preditoras (<i>input</i>) dos algoritmos de IA	24h	2000 a 2019 das 1200Z	102.143	Universidade do Wyoming (wyoweb.uwyo.edu/)
DA	Definição de TSC (<i>output</i>)	300ns	Dez/2000 a Jul/2017	6.576.737	RINDAT (www.inpe.br/webelat/rindat/)

4.2.1 DADOS DE DESCARGAS ATMOSFÉRICAS

Para esta pesquisa, foram utilizados o total de registros de DA (do tipo nuvem solo) ocorridos na TMA-SP no período de 2000 a 2019. Esses dados consistem em latitude, longitude, data, hora e minuto de eventos oriundos dos sensores da Rede Integrada Nacional de Detecção de Descargas Atmosféricas (RINDAT).

A RINDAT é constituída por uma rede de sensores que detecta em tempo real as ocorrências de DA da nuvem ao solo em parte do território brasileiro. Sua origem se deu por intermédio de um convênio de cooperação técnico-científico entre quatro instituições: Instituto Nacional de Pesquisas Espaciais (INPE), Companhia Paranaense de Energia (COPEL), através do Sistema Meteorológico do Paraná (SIMEPAR), Companhia Energética de Minas Gerais (CEMIG) e FURNAS Centrais Elétricas S.A (ELETROBRÁS FURNAS).

O sistema possui uma resolução temporal de 300 ns e espacial de aproximadamente 500 m. A RINDAT é formada por 4 Centros Operacionais localizados em Belo Horizonte, Curitiba, Rio de Janeiro e São José dos Campos. Segundo a RINDAT, a rede apresenta a eficiência de detecção entre 70% e 90%, precisão de localização média entre 0,5 km e 2 km (JOURDAN, 2017).

4.2.2 ÍNDICES DE INSTABILIDADE TERMODINÂMICOS

Os IIT foram determinados utilizando os dados do perfil atmosférico (radiossondagem) das 12 h (UTC) da estação meteorológica de altitude do aeroporto de Campo de Marte, localizado na cidade de São Paulo – SP, na latitude 23° 30' 24" S e longitude 46° 38' 02" (ver Anexo II).

4.3 MÉTODO

Na Figura 4.2 é apresentado o diagrama de blocos que representa os passos idealizados e descritos para atingir o objetivo definido.

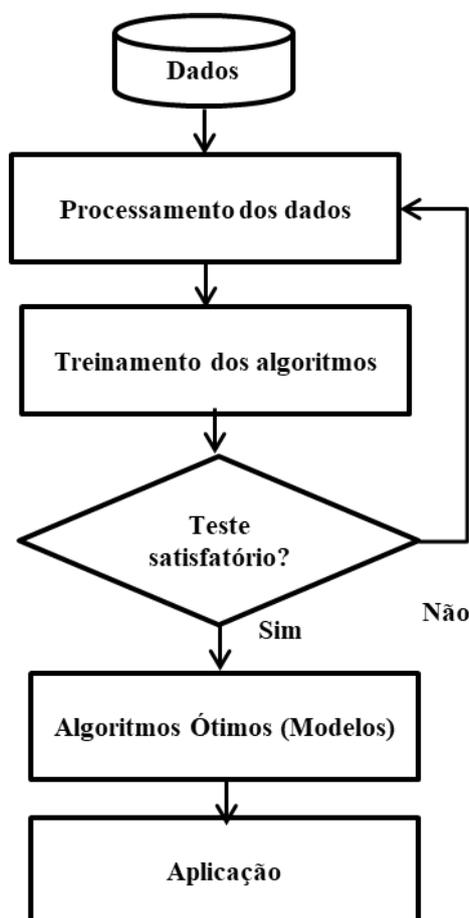


Figura 4.2: Fluxograma do método utilizado na pesquisa. Fonte: Autor.

4.3.1 DETALHES DO MÉTODO

Assumindo que os IIT foram idealizados para representar o padrão termodinâmico de uma porção atmosfera, representada pelo volume cilindro de raio de aproximadamente 150 km (WMO, 2010), e, que nesta pesquisa, pretende-se classificar, via algoritmos de AM, o padrão atmosférico que antecede o surgimento de eventos meteorológicos associados às ocorrências de TSC, foram realizados os seguintes passos:

- I. **Dados** - analisou-se a consistência dos dados históricos das radiossondagens e dos IIT;
- II. **Período de previsão** - analisou-se o comportamento das descargas atmosféricas e definiu-se o período de previsão do evento convectivo considerando o início da previsão às 10 h (horário local);
- III. **Seleção de *input* e *output*** - analisou-se correlação cruzada dos IIT disponíveis, conforme Anexo II e removeu-se os IIT redundantes. Por fim, conectou-se, cronologicamente, o conjunto de IIT e o total de ocorrências de DA diária. Os IIT selecionados foram definidos como *input* (variáveis preditivas) e o *output*, se SIM, caso existisse ocorrência de DA (no período da previsão), ou o NÃO, caso contrário;
- IV. **Modelos ótimos para previsão de evento convectivo** - estes foram obtidos via treinamento e teste dos algoritmos de AM disponíveis no WEKA (versão 3.8.5), com *input* e *output* definidos no passo anterior, até a obtenção dos modelos ótimos, via correlação cruzada⁴, observando os valores das métricas estatísticas definida na Seção 4.3.3, quando as estatísticas de Probabilidade de Detecção do Evento (POD) e Falso Alarme (FAR) atendessem a condição de $POD \geq 0,80$ e $FAR \leq 0,15$ (ou $1 - FAR \geq 0,85$);

⁴ Procedimento que divide o conjunto de dados, em k subconjuntos mutuamente exclusivos do mesmo tamanho, e cada subconjunto é utilizado para teste e os k-1 restantes são utilizados para treinamento. Este procedimento é executado k vezes alternando o subconjunto de testes até que todos os subconjuntos sejam utilizados para treinamento e avaliação de teste.

- V. **Classificação da severidade do evento convectivo** - analisou-se a distribuição de valores de cada IIT selecionado versus 12 intervalos de DA (ou seja, sem DA, 1-10, 11-100, 101-500, 501-1000, 1001-2000, 2001-3000, 3001-4000, 4001-5000, 5001-6000, 6000-10.000 e > 10.001) e definiu-se uma nova variável preditiva, denominada de Índice de Severidade (IS, eq. 4.1), como sendo a média ponderada dos valores da mediana dos IIT selecionados. Os pesos ótimos (P_i) de cada IIT_i foram obtidos por tentativa e erro variando o peso de 0,1 até K, em intervalo de 0,1, observando os resultados dos modelos ótimos;
- VI. **Aplicação do modelo** – consistiu-se em realizar *handcast* de eventos considerando uma amostra independente de eventos convectivos e não convectivos.

$$IS = \frac{\sum_{i=1}^n P_i IIT_i}{\sum P_i}; \quad (\text{Eq 4.1})$$

4.3.3 ESTATÍSTICAS DE AVALIAÇÃO

Para a avaliação do desempenho dos algoritmos classificatórios foram utilizadas as métricas estatísticas abaixo que são baseadas na matriz de confusão, conforme descrita na Tabela 4.2.

Tabela 4.2: Matriz de Confusão. Fonte: (WILKS, 2006).

		Verdade	
		Condição Positiva	Condição Negativa
Prognóstico	Verdadeiro Positivo (a)		Falso Positivo (c)
	Falso Negativo (b)		Verdadeiro Negativo (d)

Para a validação dos testes dos modelos, foram utilizados os seguintes índices estatísticos baseados na tabela de contingência:

- POD: Probabilidade de Detecção do Evento; fornece a taxa de previsão correta do evento desejado, expresso pela Equação 4.2.

$$POD = \frac{a}{a + c}; \quad (\text{Eq. 4.2})$$

- FAR: Probabilidade de Falsa Detecção ou alarme falso (ou falso positivo). Corresponde à razão do número de alarme falso dividido pelo total de não eventos observados, expresso pela Equação 4.3. Nesta pesquisa, foi utilizado 1-FAR como índice para verificação da eficiência das previsões, considerado o valor 1 como ideal.

$$FAR = \frac{b}{b + d}; \quad (\text{Eq. 4.3})$$

- F-MEASURE: É uma medida harmônica entre a precisão e a sensibilidade, expresso pela Equação 4.4.

$$F - MEASURE = \frac{2P \cdot POD}{P + POD}; \quad (\text{Eq. 4.4})$$

onde, $p = \frac{a}{a+b}$

- KAPPA: É uma medida de concordância entre o previsto e o observado, utilizado para medir o desempenho de um algoritmo de classificação, expresso pela fórmula

$$k = \frac{p_o - p_e}{1 - p_e}; \quad (\text{Eq. 4.5})$$

onde, $p_o = \frac{a+d}{a+b+c+d}$ e $p_e = \frac{[(a+b)(a+c)] + [(c+d)(b+d)]}{(a+b+c+d)^2}$.

De acordo com Landis (1977), os intervalos do índice Kappa podem ser interpretados conforme Tabela 4.3.

Tabela 4.3: Interpretação para os valores do índice Kappa (LANDIS, 1977).

Valor do Índice Kappa	Concordância
≤ 0	Pobre/Insignificante
0 – 0,20	Fraca
0,21 – 0,40	Considerável

0,41 – 0,60	Moderada
0,61 – 0,80	Substancial/Forte
0,81 – 1	Excelente/Quase perfeita

- BIAS: Mede a proporção da frequência dos eventos preditivos pela frequência dos eventos observados, utilizado para indicar se o prognóstico está superestimado ou subestimado. O valor ideal esperado é 1. O índice BIAS pode ser expresso pela Equação 4.6.

$$B = \frac{a + b}{a + c}; \quad (\text{Eq. 4.6})$$

5. RESULTADOS

Neste capítulo são discutidos o comportamento dos IIT versus a severidade dos eventos convectivos e a determinação dos modelos ótimos (definidos como algoritmos testados baseados em AM) para a previsão de eventos convectivos na TMA-SP.

5.1 ANÁLISE DOS DADOS

Esta seção trata do processamento dos dados para a aplicação nos treinamentos dos algoritmos de aprendizado de dados na construção dos modelos preditivos de eventos convectivos.

5.1.1 DADOS DE RADIOSSONDAGENS

Os dados de radiossondagens foram obtidos conjuntamente com vinte e três IIT no endereço <http://wyoweb.uwyo.edu/>, conforme Anexo II. Os índices foram armazenados e organizados cronologicamente por data e hora em uma base de dados. Seguindo os passos do método, realizou-se a correlação cruzada dos IIT para remoção dos mais correlacionados, e assim se reduziu de vinte e três para apenas seis IIT, ou seja, SH, LI, SW, K, TT e AP que são detalhados na Seção 3.3. Os IIT selecionados são as variáveis preditiva (*input*) a serem utilizadas pelos algoritmos de AM nos passos de treinamento-teste para identificar o ambiente atmosférico propenso ao surgimento ou não de evento convectivo.

Na Figura 5.1, é apresentado o número de sondagem efetivamente realizadas em SBMT às 12h (UTC) por ano, para o período de dados. Observa-se, excluindo o ano de 2000, um expressivo percentual de ausência de dados de radiossondagens, cuja média, mínimo e máximo percentual de ausência são, respectivamente iguais a 24.1%, 19.2% e 31.5%.

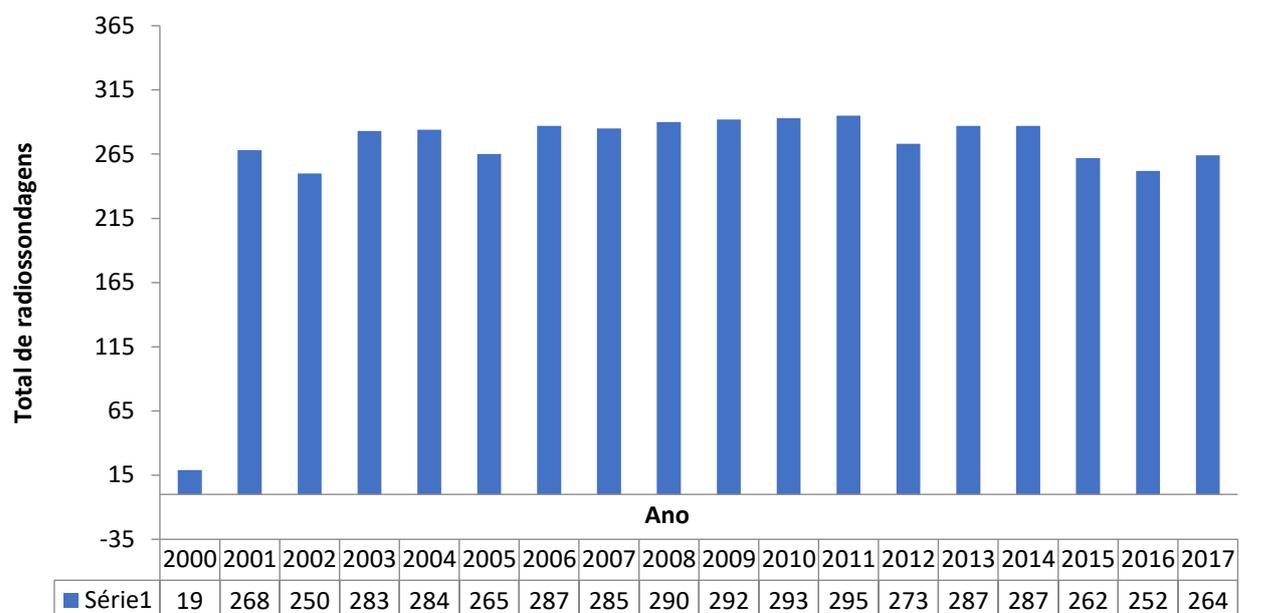


Figura 5.1: Total de radiosondas por ano disponíveis para esta pesquisa. Fonte: Autor, 2020.

5.1.2 DESCARGAS ATMOSFÉRICAS

A Tabela 5.1 sumariza 6.576.737 de DA e sua distribuição mensal das ocorrências dentro da área correspondente ao círculo de raios de 137 Km, centrada no Aeroporto de Marte, no período de dezembro de 2000 a junho de 2017. Observar-se que há diversos meses sem registros de DA e que aqui foram excluídos da amostra de dados para análise.

Na Figura 5.2 é apresentada a densidade de ocorrências de DA e observa-se que as maiores concentrações ocorrem sobre as ilhas de calor localizadas nas cidades de São Paulo (representados pelos aeródromos identificados por: SBGR, SBSP e SBMT, SBKP, SBSJ, na Serra da Mantiqueira e ao longo da Rodovia Dutra, localizadas, respectivamente, à norte e à leste de São José dos Campos.

A Figura 5.3 (a) e (b) mostram, respectivamente, o percentual da distribuição de ocorrência mensal e a quantidade de DA horária para período. Nota-se que 80% das ocorrências DA estão concentradas nos meses de verão (dezembro-março) e 84% no período entre 15 até 00h (local). Assim, considerando o passo II do método, o período de previsão aqui definido é de 5 h, correspondendo ao intervalo que inicia às 10h (local), quando dados da radiossondagens de SBMT estarão disponíveis para cálculo dos IIT e para a

execução do modelo preditivo, e às 15 h (local), quando inicia o período de maior ocorrência diária dos eventos estudados.

Tabela 5.1: Ocorrência mensal de DA na área da pesquisa do período de 2000 a 2017.

	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
2000	0	0	0	0	0	0	0	0	0	0	0	32.461
2001	126.227	102.110	84.100	5.599	1.792	1.584	817	5.975	2.789	21.775	23.738	18.953
2002	72.602	47.420	41.426	10.132	675	0	52	810	2.064	18.843	31.404	20.344
2003	81.230	84.756	58.461	2.406	768	666	786	204	188	13.358	28.985	106.971
2004	82.222	95.190	71.736	75.359	25.315	2	4.492	88	10.521	24.338	31.928	16.333
2005	75.647	43.021	31.220	14.267	0	124	1.444	8.062	5.061	56.268	9.692	64.867
2006	119.690	118.136	82.950	7.834	44	298	1.338	438	5.961	3.934	54.836	37.239
2007	29.547	82.399	121.860	34.803	62	51	5.540	14	6.772	42.040	18.524	23.771
2008	135.746	154.282	83.612	54.816	81	310	8	193	1.978	46.867	42.280	58.333
2009	87.476	146.382	45.481	10.684	2.456	5.568	11.290	2.584	14.000	19.186	84.305	90.896
2010	295.379	112.467	129.785	4.539	883	0	3.952	4	15.214	6.332	11.094	16.893
2011	123.001	105.345	10.028	15.620	858	1.280	0	750	56	3.049	19.829	20.104
2012	64.052	78.014	40.992	11.514	200	5.339	3.023	51	981	18.088	8.042	140.049
2013	88.647	224.576	141.747	3.585	2.193	6.025	1.609	2	7.094	28.582	17.646	20.862
2014	228.105	101.182	82.872	8.235	5.135	391	260	3.382	15.083	15.643	16.742	43.539
2015	0	10.358	486	0	0	0	0	0	8.705	21.582	23.240	41.552
2016	51.431	148.790	112.814	329	5.747	11	1	980	4.330	14.484	7.191	0
2017	28.097	21.501	44.006	1.415	3.073	104	0	0	0	0	0	0
Total	1.689.099	1.675.929	1.183.576	261.137	49.282	21.753	34.612	23.537	100.797	354.369	429.476	753.167

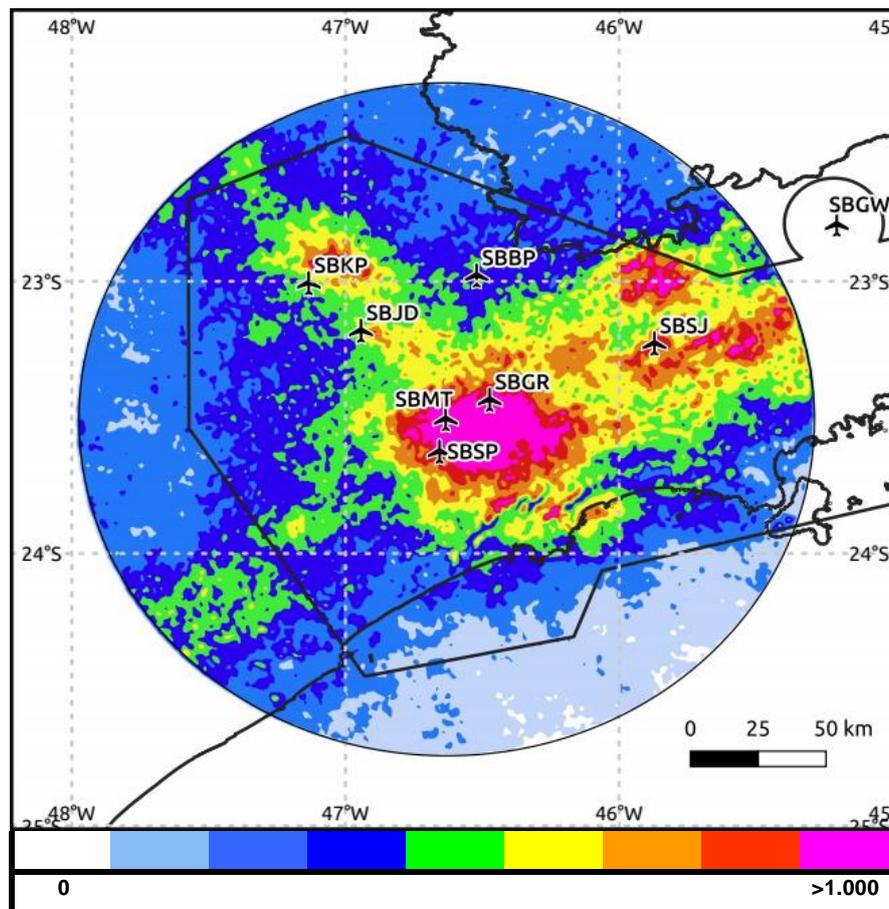
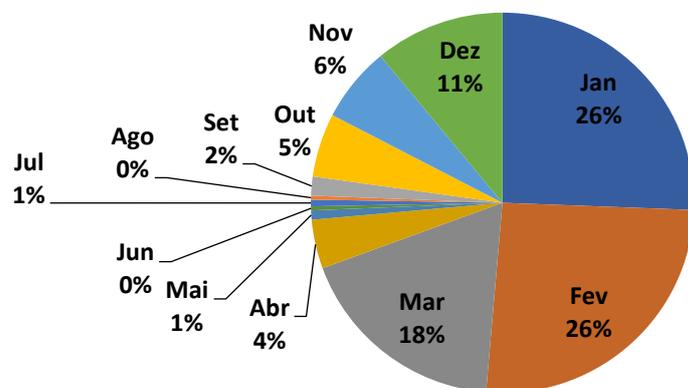
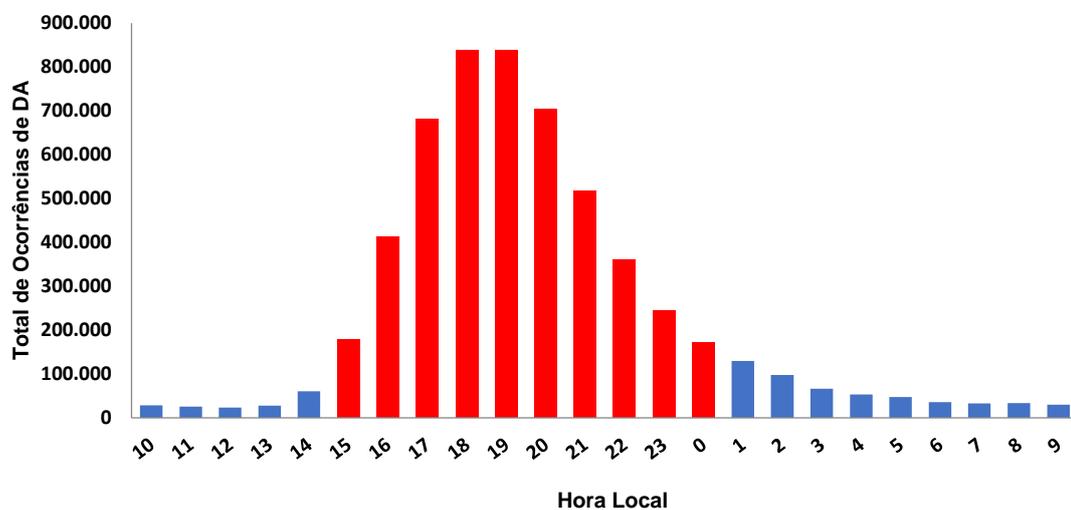


Figura 5.2: Densidade de ocorrências de DA no período de 2000 a 2017. Fonte: Autor, 2020.



(a)



(b)

Figura 5.3: Distribuição mensal (a) e horária (b) de ocorrências de DA no período de dezembro de 2000 a junho de 2017. Fonte: Autor, 2020.

5.3 PREVISÃO DE EVENTO CONVECTIVO

É sabido que o processo de treinar e testar algoritmos de AM é complexo e consome tempo. O êxito deste procedimento está associado, principalmente, na escolha das variáveis (*input*), que aqui caracterizam o estado termodinâmico local da atmosfera que precede a ocorrência dos eventos convectivos na área de estudo. A exemplo de Freitas *et al.* (2019) e Almeida *et al.* (2020), os índices SH, LI, SW, K, TT e AP foram aqui selecionados como acima mencionado, como *input* ou variáveis preditoras dos eventos convectivos (ou a não ocorrência), caracterizados pela ocorrência de DA entre 15h e 00h (local).

Conforme descrito no passo IV do método, realizou-se 384 experimentos de treinamento e teste usando todo conjunto de dados e aplicando os vinte quatro algoritmos do WEKA descritos no Anexo I, para previsão para 5 horas de evento convectivo cujo número de DA é superior a um dos dezesseis limites definidos na coluna 2 da Tabela 5.2.

Tabela 5.2: Valores das estatísticas dos melhores resultados iniciais da previsão-5h(15-00h) de evento convectivo (ou não, e sua severidade (quantidade de DA)).

No do Experimento	Quantidade de DA	Modelo	POD	1-FAR	BIAS	Kappa	F-Measure
1	0	RandomForest	0,91	0,95	0,92	0,74	0,88
2	>10	RandomForest	0,94	0,88	0,91	0,83	0,84
3	>100	RandomForest	0,83	0,85	0,97	0,74	0,84
4	>200	BayesNet	0,74	0,91	0,82	0,71	0,82
5	>300	BayesNet	0,76	0,80	0,96	0,68	0,78
6	>400	BayesNet	0,68	0,92	0,74	0,67	0,78
7	>500	RandomForest	0,75	0,67	1,12	0,62	0,71
8	>600	BayesNet	0,67	0,77	0,88	0,61	0,72
9	>700	LMT	0,60	0,91	0,66	0,62	0,73
10	>800	AdaBoostM1	0,63	0,77	0,82	0,60	0,69
11	>900	AdaBoostM1	0,58	0,83	0,70	0,58	0,69
12	>1000	Multilayer	0,70	0,56	1,25	0,55	0,63
		Perceptron					
13	>2000	Multilayer	0,66	0,71	0,94	0,60	0,68
		Perceptron					
14	>3000	JRip	0,56	0,49	1,15	0,45	0,52
15	>4000	BayesNet	0,26	0,84	0,31	0,31	0,40
16	>5000	BayesNet	0,21	0,83	0,26	0,27	0,34

A Tabela 5.2 apresenta as estatísticas dos melhores resultados de previsões-5h produzidos com os modelos (coluna 3) para dezesseis intervalos de DA, considerando o SIM como evento com DA > limite da coluna 2, caso contrário, considerando NÃO. Os treinamento e teste para modelos ótimos foram obtidos via correlação cruzada conforme descrito no passo IV. Assumindo que um modelo é ótimo quando seus resultados produzirem $POD \geq$

0,8 e $1-FAR \geq 0,85$, observa-se que apenas os resultados dos experimentos de números 1, 2, e 3 são satisfatórios para realizar a previsão-5 dos eventos convectivos até o limite de $DA=100$ e, isso, em outras palavras, demonstram que os modelos ótimos, neste conjunto de experimentos, não são capazes para indicar se o evento convectivo é ou não severo ou impactante à aviação.

5.4 CLASSIFICAÇÃO DA SEVERIDADE DO EVENTO CONVECTIVO

Considerando que os resultados das previsões para 5 hora dos eventos convectivos severos na TMA-SP são insatisfatórios, com modelos ótimos acima discutidos e, assim, na tentativa de aprimorar a previsão da severidade dos eventos, definiu-se uma nova variável preditiva que consiste na média ponderada dos valores da mediana dos IIT selecionados, que é denominada IS (Eq. 4.7), descrita no passo V do método. Sendo assim, analisou-se o comportamento médio dos IIT versus doze intervalos de DA (ou seja, sem DA, 1-10, 11-100, 101-500, 501-1000, 1001-2000, 2001-3000, 3001-4000, 4001-5000, 5001-6000, 6000-10.000 e > 10.001), conforme apresentado na Figura 5.4. Observa que o comportamento dos IIT (SH, LI, SW, K, TT e AP) é correlacionado aos acumulados diários de DA e possuem valores absolutos da correlação variando entre a 0,96% e 0,99% e que a medida que os eventos convectivos ficam mais severos, a partir, aproximadamente, do limite $DA \geq 1000$, os valores dos índices convergem para a mediana (dentro dos intervalos estreitos do 1º ao 3º quartis da distribuição) que, em outras palavras, quer dizer que se torna complexo, a princípio, via algoritmos de análise de agrupamento, que utilizam métricas quantitativas tradicionais, classificar eventos convectivos com a quantidade de DA superior a mencionada. Desta forma, o uso de algoritmos classificatórios de AM é sugestivo, conforme recomendado por Almeida et al. (2020).

A análise da Figura 5.4 revela que os índices SH, LI e AP (definido como GRUPO 1) são mais determinantes que os índices SW, K, TT (definido como GRUPO 2) para a classificação do limite severidade de $DA \geq 1000$.

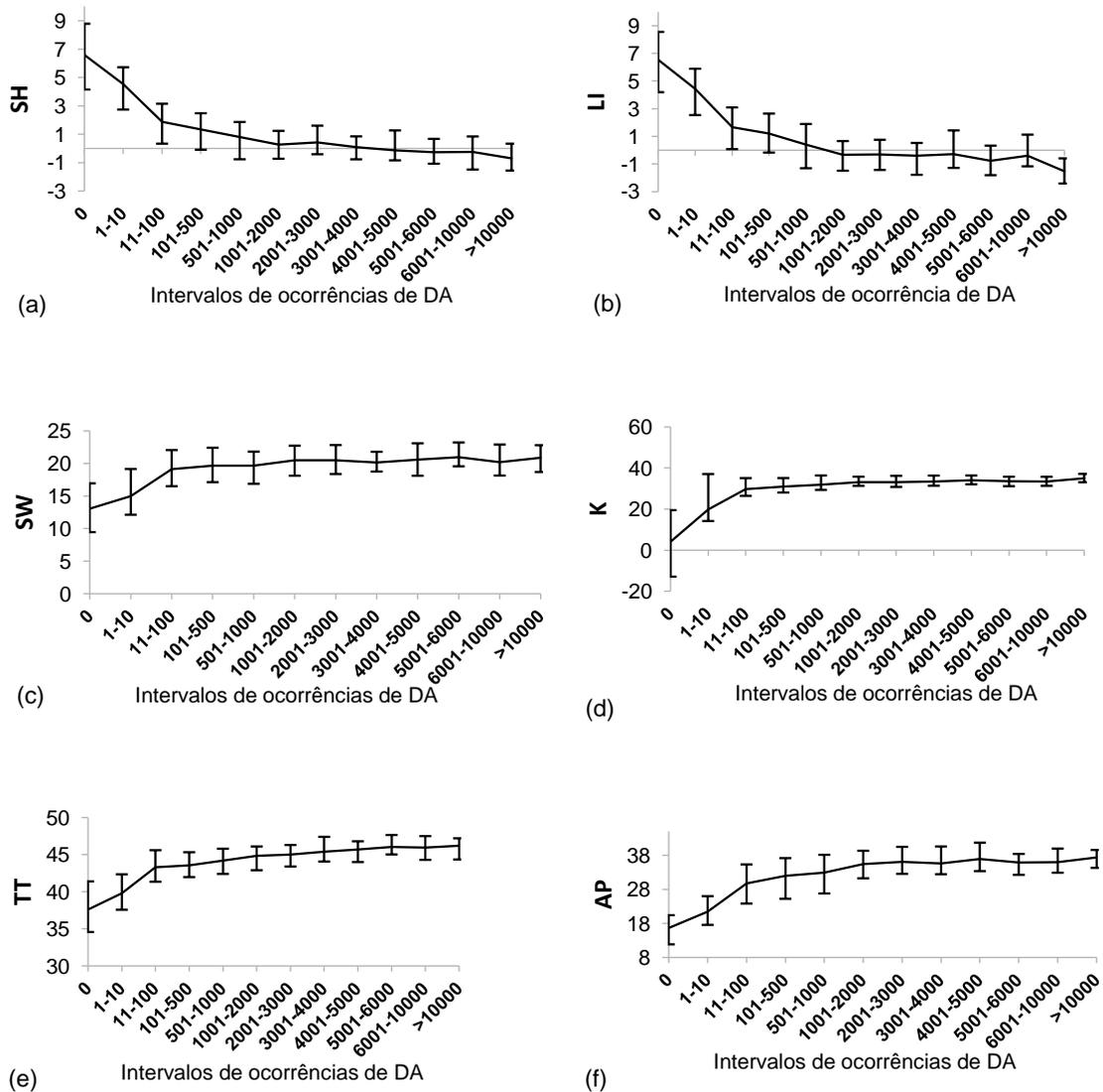


Figura 5.4: Comportamento dos índices atmosféricos SH (a), LI (b), SW (c), K (d), TT (e) e AP (f) versus o intervalo do total de ocorrências de DA. Fonte: Autor, 2020.

Para determinação dos pesos de cada IIT da Eq. 4.7, define-se evento convectivo severo aqueles com ocorrências de DA ≥ 1000 . Seguindo o descrito no passo V do método, os pesos ótimos (P_i) para cada IIT $_i$ são obtidos por tentativa e erro variando o P_i de 0,1 até $k=3$, em intervalo de 0,1, observando os resultados dos algoritmos quanto a valores de POD e FAR ótimos definidos. Este procedimento resultou em pesos para os índices do GRUPO 1 e GRUPO 2 respectivamente iguais a 2 e 1. Desta forma, o IS é incorporado, conjuntamente aos IIT selecionados, como variáveis preditivas no processamento de treinamento e testes dos vinte e quatro algoritmos.

A exemplo dos resultados apresentados na Tabela 5.2, a Tabela 5.3 apresenta as estatísticas dos melhores resultados dos modelos de previsão para 5 horas de eventos convectivos para TMA-SP considerando os resultados de 38.401 experimentos realizados considerando 16 intervalos de DA. Nota-se que a previsão para 5 horas de eventos convectivos até o limite de $DA \geq 1000$ obteve significativos índices de acerto, visto que os valores da POD, 1-FAR, BIAS, KAPPA e F-MEASURE são expressivos e respectivamente iguais 0,85, 0,94, 0,89, 0,77, e 0,95. Isto significa, em outras palavras, que a inclusão do IS, como variável preditiva, aprimorou, significativamente, a previsão para 5 horas da severidade dos eventos convectivos, visto que estas são capazes de discriminar com precisão a severidade do evento convectivo até de 1000 ocorrências de DA.

Analisando os valores de KAPPA e as previsões para 5 horas de eventos sem DA e $DA \leq 1000$, observa-se que $Kappa \geq 0,6$, o que indica, conforme Tabela 4.3, uma concordância substancial nos prognósticos. Ainda, para os experimentos com o prognóstico da ocorrência do evento, considerando a ausência de ocorrência de $DA \leq 300$, os resultados do índice Kappa obtiveram resultados com valores superiores a 0,81 o que representa uma concordância ótima nos prognósticos, conforme Tabela 4.3.

Tabela 5.3: Resultado dos Modelos Preditos.

No do Experimento	Quantidade de DA	Modelo	POD	1-FAR	BIAS	Kappa	F-Measure
1	0	<i>DecisionTable</i>	0,92	0,91	1,00	0,83	0,90
2	>10	<i>RandomForest</i>	0,93	0,97	0,96	0,89	0,95
3	>100	<i>RandomForest</i>	0,91	0,93	0,98	0,84	0,94
4	>200	<i>RandomForest</i>	0,90	0,93	0,97	0,82	0,94
5	>300	<i>LMT</i>	0,92	0,93	0,99	0,83	0,95
6	>400	<i>LWL</i>	0,96	0,90	1,07	0,80	0,94
7	>500	<i>LogitBoost</i>	0,88	0,92	0,95	0,79	0,94
8	>600	<i>DecisionTable</i>	0,84	0,93	0,90	0,76	0,94
9	>700	<i>RandomForest</i>	0,85	0,92	0,91	0,75	0,94
10	>800	<i>J48</i>	0,84	0,92	0,91	0,73	0,94
11	>900	<i>LMT</i>	0,82	0,94	0,86	0,74	0,95
12	>1000	<i>REPTree</i>	0,85	0,94	0,89	0,77	0,95
13	>2000	<i>Jrip</i>	0,73	0,92	0,77	0,59	0,94
14	>3000	<i>Jrip</i>	0,68	0,97	0,69	0,60	0,97
15	>4000	<i>MultilayerPerceptron</i>	0,66	0,95	0,68	0,52	0,97
16	>5000	<i>Kstar</i>	0,63	0,94	0,65	0,43	0,97

5.4 APLICAÇÃO: ESTUDO DE CASOS

A seção discorre sobre aplicação dos modelos ótimos obtidos em trinta eventos meteorológicos aleatórios selecionados como estudo de casos.

A Tabela 5.4 apresenta o resultado do modelo ótimo, utilizando os valores IIT das 12h (UTC), que conjuntamente com IS foram usados com preditores (*input*) para realizar o *handcast* para previsões de 5 horas (*handcast-5h*) na TMA-SP dos 30 (trinta) casos especificados na coluna 2. Observa-se que os modelos foram capazes de realizar o *handcast-5h* com alta precisão, classificando 100% do evento meteorológico não convectivo; 97% daqueles convectivos e 87% de acerto se os mesmos são ou não severos.

A seguir é apresentado uma discussão dos aspectos sinóticos para os casos em que os *handcasts* dos modelos utilizados não obtiveram resultados satisfatórios.

Tabela 5.4: Resultado do Estudo de caso com 30 amostras aleatórias.

Eventos	Dia	DA	Resultado dos Modelos Preditivos (ACERTO: A – ERRO: E)	
			Detecção do Evento	Severidade
1	10/08/2004	0	A	A
2	17/01/2006	208	E	E
3	09/07/2007	0	A	A
4	12/09/2007	0	A	A
5	14/08/2013	0	A	A
6	22/02/2013	12950	A	E
7	05/03/2013	22796	A	A
8	17/04/2005	199	A	A
9	15/04/2002	106	A	A
10	28/12/2003	609	A	A
11	26/01/2003	473	A	A
12	23/02/2005	620	A	A
13	24/12/2002	2865	A	A
14	29/11/2013	1640	A	A
15	16/01/2005	3361	A	A
16	07/02/2006	7103	A	A
17	17/02/2011	3619	A	A
18	21/04/2004	2190	A	A
19	03/03/2006	4844	A	A
20	23/03/2006	1843	A	A
21	28/11/2001	8095	A	A
22	22/02/2017	1370	A	A
23	11/01/2012	18	A	E
24	19/02/2013	14715	A	A
25	17/01/2010	2763	A	A
26	16/01/2014	17556	A	A
27	13/03/2013	77	A	E
28	14/02/2008	13497	A	A
29	09/01/2016	6933	A	A
30	15/02/2016	16398	A	A

5.4.1 ANÁLISE DO DIA 22 DE FEVEREIRO DE 2013

Dentre os dias selecionados no *handcast* apresentado na Tabela 5.4, foi realizado uma breve análise de dois erros obtidos pelo modelo, no prognóstico de severidade, no qual, o primeiro, do dia 22 de fevereiro de 2013, onde todos os IIT apontavam para uma atmosfera estável e sem previsão de ocorrência de TSC, sendo registrado 12.950 ocorrências de DA.

Na Figura 5.5 (a), pode-se verificar a passagem de um sistema frontal próximo da área estudada, porém, no momento da radiossondagem, a atmosfera apresentava-se ainda estável e, na Figura 5.5 (b), verifica-se a atuação do sistema frontal (na área estudada) com muitas nuvens convectivas o que evidencia a ocorrência de TSC.

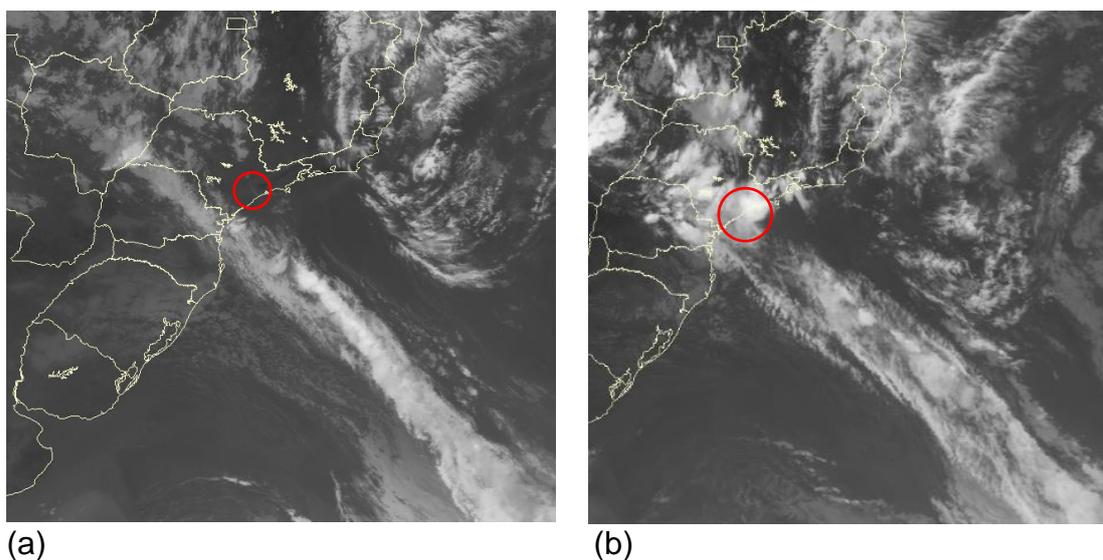


Figura 5.5: Imagens de Satélite GOES-13, do canal Infravermelho (CH04), do dia 22 de fevereiro de 2013 das 1230Z (a) e 2000Z (b). Fonte: Sítio de Internet do CPTEC/INPE, acessado em 06 de maio de 2021.

5.4.2 ANÁLISE DO DIA 13 DE MARÇO DE 2013

Ainda, como análise de erro do modelo, foi selecionado, conforme a Tabela 5.5, o dia 13 de março de 2013, onde todos os índices apontavam para uma atmosfera muito instável, com previsão de ocorrência de TSC.

Conforme Figura 5.6 (a), verifica-se que, durante a realização da radiossondagem (1200Z), atuava, na área de estudo, um sistema frontal apresentando uma atmosfera instável e, na Figura 5.6 (b), verifica-se que, às 1800Z, a rampa frontal associada ao sistema, já se encontrava no oceano e, desta forma, na área estudada não se observa muita instabilidade, justificando, assim, a presença do sistema frontal para o erro do modelo.

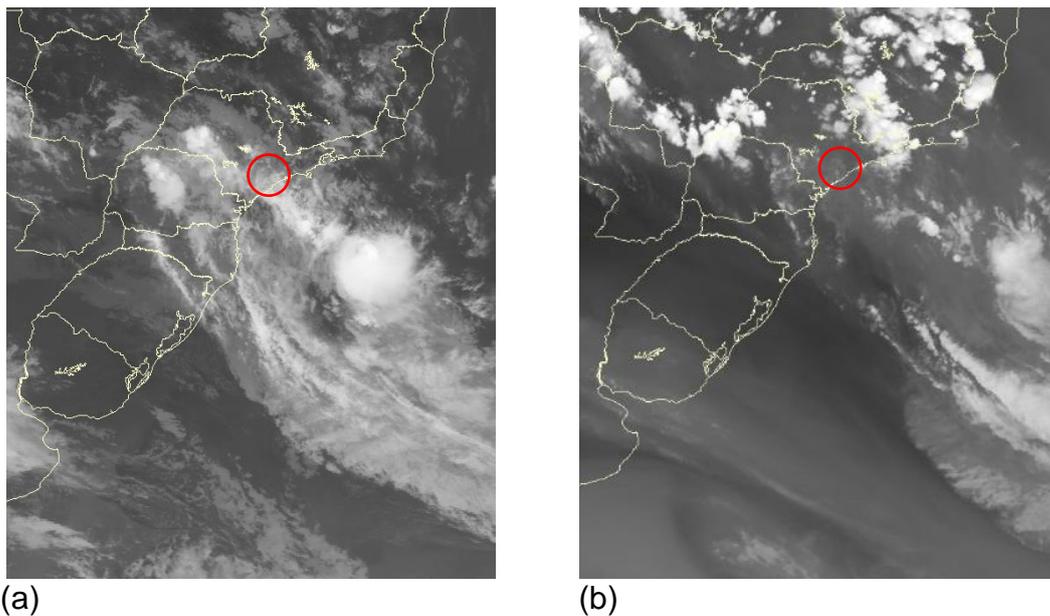


Figura 5.6: Imagens de Satélite GOES-13, do canal Infravermelho (CH04), do dia 22 de fevereiro de 2013 das 1230Z (a) e 1800Z (b). Fonte: Sítio de Internet do CPTEC/INPE, acessado em 06 de maio de 2021.

Em geral, os modelos são eficazes em prever os eventos convectivos locais e sua severidade, errando em apenas dois casos quando a quantidade de DA dos eventos foram inferiores a 100 ocorrências de DA, em 11 de janeiro de 2012 e 13 de março de 2013. Além disso, os resultados também demonstram que ainda há dificuldade dos modelos na previsão de evento severo convectivo devido entrada de sistema frontal no verão (quando há mais disponibilidade de energia que potencializa a convecção) na TMA-SP, como em nos dias 17 de janeiro de 2006 e 22 de fevereiro de 2013.

6. CONCLUSÃO

Esta pesquisa desenvolveu um modelo de previsão, baseado em IA visando a detecção e severidade de TSC para a TMA-SP. Os principais resultados obtidos são resumidos, a seguir:

- As descargas atmosféricas ocorrem em maior frequência no verão e são associadas, preferencialmente às ilhas de calor localizadas nas cidades de São Paulo (representados pelos aeródromos identificados por: SBGR, SBSP e SBMT), Campinas (SBKP), São José dos Campos (SBSJ), na Serra da Mantiqueira e ao longo da Rodovia Dutra, localizadas, respectivamente, à norte e à leste de São José dos Campos (SBSJ);
- Foi identificado o período das 15-00h (local) como o de maior ocorrência de atividade convectiva;
- Os modelos desenvolvidos apresentaram capacidade de prognosticar e classificar a severidade (com até 12h de antecedência), com alta probabilidade de detecção, baixa taxa de falso alarme e pouco enviesado;
- Os resultados de *handcasts* dos modelos apresentaram desempenhos satisfatórios para identificarem a ocorrência ou não de eventos convectivos e TSC para área estudada; e
- Por intermédio das análises apresentadas, pode-se concluir que os modelos não têm um bom comportamento na atuação de sistemas frontais na área estudada, devendo-se, desta forma, aplicá-los em casos em que não há influência de tais sistemas.

Como trabalhos futuros, sugere-se ajuste dos modelos para previsão eventos convectivos com origem frontal ao longo do verão e o desenvolvimento de procedimento de controle de qualidade dos dados para qualquer ação visando aplicação operacional dos modelos aqui desenvolvidos.

8. REFERÊNCIAS

ALMEIDA, V. A., FRANÇA, G. B., and Velho, H. F. C., 2020. Short-range forecasting system for meteorological convective events in Rio de Janeiro using remote sensing of atmospheric discharges, *International Journal of Remote Sensing*, 41:11, 4372-4388, DOI: 10.1080/01431161.2020.1717669.

BATTAN, L. J., 1953: Duration of convective radar cloud units. *Bull. Amer. Meteor. Soc.*, 34, 227-228.

COHEN, J. A. 1960. Coeficient of Agreement for Nominal Scales. *Educational and Measurment*. v. XX, n. 1, p. 37-46.

DECEA. Anuário Estatístico de Tráfego Aéreo. Departamento de Controle do Espaço Aéreo, Rio de Janeiro, Brasil, 2019.

ELAT. Grupo de Eletricidade Atmosférica. Disponível no site da Instituto Nacional de Pesquisas Espaciais (2021). URL: <http://www.inpe.br/webelat/homepage/menu/relamp/relampagos/origem.php>.

FRANÇA, G. B., de ALMEIDA, M. V., and ROSETTE, A. C.: An automated nowcasting model of significant instability events in the flight terminal area of Rio de Janeiro, Brazil, *Atmos. Meas. Tech.*, 9, 2335–2344, <https://doi.org/10.5194/amt-9-2335-2016>, 2016.

GALWAY, J. G. 1956. The lifted index as a predictor of latent instability. *Bull. Amer. Meteor. Soc.* p. 528–529.

GEORGE, J. J. 1960. *Weather Forecasting for Aeronautics*. Academic Press. p. 673.

GREWAL, D. S., 2014. A Critical Conceptual Analysis of Definitions of Artificial Intelligence as Applicable to Computer Engineering. *IOSR Journal of Computer Engineering*, v. 16, i 1, p. 09-13.

GROSSMAN, K.S. 2010. Estudo da combinação de índices de instabilidade como ferramenta de auxílio na previsão do tempo. Departamento de

Meteorologia (Instituto de Geociências). Universidade Federal do Rio de Janeiro, p. 68.

GULTEPE, I.; SHARMAN, R.; WILLIAMS, P.D.; ZHOU, B.; ELLROD, G.; MINNIS, P.; TRIER, S.; GRIFFIN, S.; YUM, S.S.; GHARABAGHI, B.; et al, 2019. A review of high impact weather for aviation meteorology. *Pure Appl. Geophys.* p. 1869–1921.

FREITAS, J. H; FRANÇA, G. B. MENEZES, W. F., 2018. Previsão de instabilidade atmosférica significativa usando árvore de decisão na região metropolitana do Rio de Janeiro. 2018. 109 f. Dissertação (Mestrado) - Curso de Meteorologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2018. HAYKIN, S. (2002). *Neural Networks. A Comprehensive Foundation*, Macmillan, New York, NY.

ICAO. *Global Air Navigation Plan*. 6ª ed., 2016. 142 f., Organização de Aviação Civil Internacional, Montreal, Canadá.

ISAAC, G. A., BURROWS, W. R., COBER, S. G., CRAWFORD, R. W., DONALDSON, N., GULTEPE, I., HANSEN, B., HECKMAN, I., HUANG, L. X., LING, A., MAILHOT, J., MILBRANDT, J. A., REID, J., and FOURNIER, M., 2014. The Canadian airport nowcasting system (CANNOW), *Meteorol. Appl.*, 21, p. 30–49.

JOHN, R. H.; DOSWELL, C. A., 1992. Severe local storms forecasting. *Weather and Forecasting*, v. 7, p. 588-612.

JOHNS, R. H.; DOSWELL, C. A. 1992. Severe local storms forecasting. *Wea. Forecasting*, v. 7, p. 588-612.

JOURDAN, P. et al, 2017. Relações Espaciais e Temporais entre o Desligamento de uma Linha de Transmissão e a Detecção de Descargas Atmosféricas em casos de ruptura do cabo OPGW. XXIV SNPTEE – Seminário Nacional de Produção e Transmissão de Energia Elétrica. Curitiba, PR.

KANE, RICHARD J., 1991. Correlating lightning to severe local storms in the northeastern United States. *Wea. Forecasting*, 6, p. 3-12.

LANDIS, J.R.; KOCH, G.G., 1977. The measurement of observer agreement for categorical data. *Biometrics*. Vol. 33, p. 159–174.

LIU, J. G.; MASON, P. J., 2016. *Image Processing and GIS for Remote Sensing: Techniques and Applications*. Wiley Blackwell.

MCCANN, D., 1992. A neural network short-term forecast of significant thunderstorms. *Wea. Forecasting*, 7, 525–534, [https://doi.org/10.1175/1520-0434\(1992\)007,0525:ANNSTF.2.0.CO;2](https://doi.org/10.1175/1520-0434(1992)007,0525:ANNSTF.2.0.CO;2).

MILLER, R. C., 1972. Notes on analysis and severe storm forecasting procedures of the Air Force Global Weather Central. Tech. Rept. 200(R). Headquarters, Air Weather Service, USAF. p. 190.

MOURA, C.R.W.; ESCOBAR, G.C.J.; ANDRADE, K.M., 2013. Padrões de circulação em superfície e altitude associados a eventos de chuva intensa na Região Metropolitana do Rio de Janeiro. *Revista Brasileira de Meteorologia*, v.28, n.3, p. 267 – 280.

MOURÃO, C. E. F.; MENEZES, W. F., 2006. “Estudo do comportamento de indicadores de tempo severo em casos de tempestades sobre o Rio de Janeiro”. *Anais do XIV Congresso Brasileiro de Meteorologia*. Florianópolis-SC.

MUELLER, C., SAXEN, T., ROBERTS, R., WILSON, J., BETANCOURT, T., DETTLING, S., OIEN, N., and YEE, J., 2003. NCAR auto-nowcast system, *Weather Forecast*, 18, p. 545–561.

MUNAKATA, TOSHINORI, 1994. *Fundamentals of the new Artificial Intelligence: neural, evolutionary, fuzzy and more*, 2nd ed. Springer-Verlag London. Cleveland State University.

NASCIMENTO, E.D.L., 2005. Previsão de tempestades severas utilizando-se parâmetros convectivos e modelos de mesoescala: uma estratégia operacional adotável no Brasil. *Revista Brasileira de Meteorologia*, 20(1), p.121-140.

NEWTON, C. W., and J. C. FANKHAUSER, 1964. On the movements of convective storms, with emphasis on size discrimination in relation to water-budget requirements. *J. Appl. Meteor.*, 3, p. 651-668.

NUNES, L. HIDALGO, 2015. *Urbanização e Desastres Naturais*. São Paulo: Oficina de Textos.

PAULUCCI, T. B., 2017. *Caracterização Espaço-Temporal de Descargas Atmosféricas e Tempestades Elétricas na Região Metropolitana do Rio de Janeiro entre 2001 e 2016*. Curso de Meteorologia, Universidade Federal do Rio de Janeiro, Monografia, p. 73.

PAULUCCI, T. B.; GUTEMBERG, F.; LIBONATI, R.; RAMOS, A. M., 2019. Long-Term Spatial– Temporal Characterization of Cloud-to-Ground Lightning in the Metropolitan Region of Rio de Janeiro. *Pure and Applied Geophysics*. 10.1007/s00024-019- 02216-1, 2019.

PEJOVIC, T., 2008. Implications of climate change for the UK aviation sector. p. 11-17. Centre for Transport Studies Department of Civil and Environmental Engineering Imperial College London, United Kingdom.

RIEHL, H., 1965. *Meteorologia Tropical*. USAID. Rio de Janeiro. p. 426.

SHOWALTER, A. K., 1947. A stability index for forecasting thunderstorms. *Bull. Amer. Meteor. Soc.* 34. p. 250– 252.

TURMAN, B.N., and R.1. TETTELBACH, 1980. Synoptic scale satellite lightning observations in conjunction with tornadoes. *Mon. Wea. Rev.*, 108, p. 1878-1882.

UNIVERSITY OF WAIKATO, 2010. WEKA 3 – Machine Learning Software in Java. Disponível no site da University of Waikato. URL: <http://www.cs.waikato.ac.nz/ml/WEKA>.

VAREJÃO-SILVA, MARIO ADELMO, 2006. *Meteorologia e Climatologia*. Recife.

VIERA, A. J.; MD; GARRETT, J. M., 2005. Understanding Interobserver Agreement: The Kappa Statistic. *Family Medicine Journal*, v. 37, n. 5, p. 360-363.

WEISMAN, M. L., and J. B. KLEMP, 1986. Characteristics of isolated convective storms. *Mesoscale Meteorology and Forecasting*, P. S. Ray, Ed., Amer. Meteor. Soc., p. 331-358.

WILKS, S. D., 2006. *Statistical Methods in the Atmospheric Sciences*. 2^a ed., New York, USA, Academic Press.

WILLIAMS, E.R., M.E. WEBER, and R.E. ORVILLE, 1989: The relationship between lightning type and convective state of thunderclouds. *J. Geophys. Res.*, 94, p. 13213-13220.

WILSON, J. W., 1966. Movement and predictability of radar echoes. Tech. Memo ERTM-NSSL-28, National Severe Storms Laboratory, 30 pp. [Available from National Information Service, Operations Division, Springfield, VA 22161.]

WMO. Guide to the GCOS Surface Network (GSN) and GCOS Upper-Air Network (GUAN), 2010. 34 f. World Meteorological Organization, Geneva, Suíça, 2010.

YAO, X. and LIU, Y., 2013. Machine Learning. School of Computer Science, University of Birmingham, UK, C. 12, p. 341-343.

ZHANG, H., 2004. The Optimality of Naive Bayes. American Association for Artificial Intelligence.

ANEXO I

LISTA DE ALGORITMOS UTILIZADOS NA PESQUISA⁵

ALGORÍTIMOS DO WEKA	DESCRIÇÃO DO ALGORÍTIMO	PARÂMETROS
bayes.BayesNet	Algoritmo de classificação com a aprendizagem sendo realizada utilizando vários algoritmos de pesquisa e medidas de qualidade. Fornece estruturas de dados (estrutura de rede, distribuições de probabilidade condicional etc.) e recursos comuns aos algoritmos de aprendizagem como K2 e B.	'-D -Q weka.classifiers.bayes.net.search.local.K2 -- -P 1 -S BAYES -E weka.classifiers.bayes.net.estimate.SimpleEstimator -- -A 0.5'
NaiveBayes	Algoritmo de classificação utilizando classes de estimadores.	-
Logistic	Classe de construção e utilização de modelo de regressão logística multinomial com estimador de crista.	-
MultilayerPerceptron	Um classificador que usa retropropagação para aprender um perceptron multicamadas para classificar instâncias. A rede pode ser construída manualmente ou configurada usando uma heurística simples. Os parâmetros de rede também podem ser monitorados e modificados durante o tempo de treinamento. Os nós nesta rede são todos sigmóides (exceto quando a classe é numérica, caso em que os nós de saída se tornam unidades lineares sem limites)	-
SGD	Implementa descida de gradiente estocástico para aprender vários modelos lineares (classe binária SVM, classe binária de regressão logística, perda quadrada, perda de Huber e regressão linear de perda insensível a ϵ). Substitui globalmente todos os valores ausentes e transforma os atributos nominais em binários. Ele também normaliza todos os atributos, portanto, os coeficientes na saída são baseados nos dados normalizados. Para atributos de classe numéricos, a função de perda ao quadrado, Huber ou ϵ -insensível deve ser usada. Não sensível a ϵ e perda de Huber podem exigir uma taxa de aprendizado muito maior.	'-F 0 -L 0.01 -R 1.0E-4 -E 500 -C 0.001 -S 1'

⁵ Dados obtidos do sítio de Internet do Manual do Software WEKA, disponível em <https://weka.sourceforge.io/doc.dev/allclasses-noframe.html>.

ALGORÍTIMOS DO WEKA	DESCRIÇÃO DO ALGORÍTIMO	PARÂMETROS
SimpleLogistic	Classificador para construção de modelos de regressão logística linear. LogitBoost com funções de regressão simples como aprendizes básicos é usado para ajustar os modelos logísticos. O número ideal de iterações LogitBoost a serem executadas é validado cruzado, o que leva à seleção automática de atributos.	'-I 0 -M 500 -H 50 -W 0.0'
SVM	Implementa o algoritmo de otimização sequencial mínima de John Platt para treinar um classificador de vetores de suporte. Essa implementação substitui globalmente todos os valores ausentes e transforma os atributos nominais em binários. Ele também normaliza todos os atributos por padrão. (Nesse caso, os coeficientes na saída são baseados nos dados normalizados, não nos dados originais - isso é importante para interpretar o classificador.)	'-C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K \\"weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007\\" -calibrator \\"weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4\\"'
IBk	Classificador de vizinhos K mais próximos. Pode-se selecionar o valor apropriado de K com base na validação cruzada. Também pode-se fazer ponderação de distâncias.	'-K 1 -W 0 -A \\"weka.core.neighboursearch.LinearNNSearch -A \\\\"weka.core.EuclideanDistance -R first-last\\\\"'
KStar	Classificador baseado em instância, ou seja, a classe de uma instância de teste é baseada na classe dessas instâncias de treinamento semelhantes a ela, conforme determinado por alguma função de similaridade. Ele difere de outras classes baseadas em instâncias porque usa uma função de distância baseada em entropia.	'-B 20 -M a'
LWL	Usa um algoritmo baseado em instância para atribuir pesos de instância que são então usados por um <i>WeightedInstancesHandler</i> especificado. Pode fazer classificação (por exemplo, usando Bayes ingênuo) ou regressão (por exemplo, usando regressão linear).	'-U 0 -K -1 -A \\"weka.core.neighboursearch.LinearNNSearch -A \\\\"weka.core.EuclideanDistance -R first-last\\\\" -W weka.classifiers.trees.DecisionStump'
AdaBoostM1	Classe para impulsionar um classificador de classe nominal usando o método Adaboost M1. Apenas problemas de classe nominal podem ser resolvidos.	'-P 100 -S 1 -I 10 -W weka.classifiers.trees.DecisionStump'
AttributeSelectedClassifier	O dimensionamento dos dados de treinamento e teste é reduzido pela seleção de atributos antes de serem transmitidos a um classificador.	'-E \\"weka.attributeSelection.CfsSubsetEval -P 1 -E 1\\" -S \\"weka.attributeSelection.BestFirst -D 1 -N 5\\" -W weka.classifiers.trees.J48 -- -C 0.25 - M 2'

ALGORÍTIMOS DO WEKA	DESCRIÇÃO DO ALGORÍTIMO	PARÂMETROS
Bagging	Classe para empacotar um classificador para reduzir a variância. Pode fazer classificação e regressão dependendo da classe base.	'-P 100 -S 1 -num-slots 1 -I 10 -W weka.classifiers.trees.REPTree -- -M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0'
LogitBoost	Classe para realização de regressão logística aditiva. Esta classe realiza a classificação usando um esquema de regressão como a classe base e pode lidar com problemas de várias classes.	'-P 100 -L -1.7976931348623157E308 -H 1.0 -Z 3.0 -O 1 -E 1 -S 1 -I 10 -W weka.classifiers.trees.DecisionStump'
RandomCommITee	Classe para construir um conjunto de classificadores de base aleatórios. Cada classificador base é construído usando um seed aleatório diferente (mas com base nos mesmos dados). A previsão final é uma média direta das previsões geradas pelos classificadores base individuais.	'-S 1 -num-slots 1 -I 10 -W weka.classifiers.trees.RandomTree -- -K 0 -M 1.0 -V 0.001 -S 1'
DecisionTable	Classe para construir e usar um classificador de maioria de tabela de decisão simples.	'-X 1 -S \"weka.attributeSelection.BestFirst -D 1 -N 5\"'
JRip	Esta classe implementa um aprendizado de regra proposicional, Poda incremental repetida para produzir redução de erros (RIPPER), que foi proposta por William W. Cohen como uma versão otimizada do IREP.	'-F 3 -N 2.0 -O 2 -S 1'
DecisionStump	Classe de construção e uso de toco de decisão. Normalmente utilizado em conjunto com um algoritmo de reforço. Faz regressão (com base no erro quadrático médio) ou classificação (com base na entropia). Os dados ausentes são tratados como um valor separado.	-
HoeffdingTree	Uma árvore Hoeffding (VFDT) é um algoritmo de indução de árvore de decisão incremental, a qualquer momento, que é capaz de aprender a partir de fluxos de dados massivos, assumindo que os exemplos de geração de distribuição não mudam com o tempo. As árvores Hoeffding exploram o fato de que uma pequena amostra pode muitas vezes ser suficiente para escolher um atributo de divisão ideal. Essa ideia é apoiada matematicamente pelo limite de Hoeffding, que quantifica o número de observações necessárias para estimar algumas estatísticas dentro de uma precisão prescrita.	'-L 2 -S 1 -E 1.0E-7 -H 0.05 -M 0.01 -G 200.0 -N 0.0'

J48

Classe para gerar uma árvore de decisão C4.5 podada ou não.

'-C 0.25 -M 2'

ALGORÍTIMOS DO WEKA**DESCRIÇÃO DO ALGORÍTIMO****PARÂMETROS**

LMT

Classificador para construir árvores modelo logístico, que são árvores de classificação com funções de regressão logística nas folhas. O algoritmo pode lidar com variáveis de destino binárias e multiclases, atributos numéricos e nominais e valores ausentes.

'-I -1 -M 15 -W 0.0'

RandomForest

Classe para construir uma floresta de árvores aleatórias.

'-P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1'

RandomTree

Classe para construção de uma árvore que considera K atributos escolhidos aleatoriamente em cada nó. Não executa poda. Também tem uma opção para permitir a estimativa de probabilidades de classe (ou média alvo no caso de regressão) com base em um conjunto de validação (*backfitting*).

'-K 0 -M 1.0 -V 0.001 -S 1'

REPTree

Constrói uma árvore de decisão ou de regressão usando ganho ou variância de informação e a poda usando erro reduzido (com *backfitting*). Classifica valores para atributos numéricos apenas uma vez. Os valores ausentes são tratados dividindo as instâncias correspondentes em partes (ou seja, como em C4.5).

'-M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0'

ANEXO II

LISTA DE ÍNDICES DE INSTABILIDADE TERMODINÂMICOS⁶

SIGLA	NOMECLATURA DOS ÍNDICES
SH	Índice de instabilidade SHOWALTER utilizado para previsão de tempestades
LI	Índice de Levantamento (<i>Lifted</i>) semelhante ao SH
LITV	Cálculo do índice LI usando a temperatura virtual
SW	SWEAT que representa o Índice de ameaça de Tempo Severo
K	Índice K aplicado para estimar a probabilidade de chuva e tempestades
CT	Cross Totals aplicado para prognosticar tempestades severas
VT	Vertical Totals aplicado para prognosticar o potencial desenvolvimento de tempestade severa
TT	Total Totals aplicado para prognosticar o potencial desenvolvimento de tempestade severa
CAPE	Energia Potencial Convectiva Disponível (J/kg)
CAPV	CAPE calculado utilizando temperatura virtual
CINS	Inibição de Convecção (J/kg)
CINV	CINS calculado utilizando temperatura virtual
EQLV	Nível de Equilíbrio (hPa)
EQTV	EQLV calculado utilizando temperatura virtual
LFCT	Nível de Convecção Livre (hPa) por comparação da temperatura entre uma parcela e a superfície
LFCV	LFCT calculado utilizando temperatura virtual
BRCH	Número de Bulk Richardson
BRCV	BRCH calculado utilizando CAPV
LCLT	Temperatura (K) no LCL, o nível de condensação em elevação, de uma média dos 500 metros mais baixos
LCLP	Pressão (hPa) no LCL, o nível de condensação em elevação, de uma média dos 500 metros mais baixos
MLTH	Camada média de mistura THTA (K)
MLMR	Camada média de mistura MIXR (g/kg)
THTK	Espessura entre 1000 mb e 500 (m)
PA	Água Precipitável (mm) em toda sondagem

⁶ Dados obtidos do sítio de Internet da Universidade do Wyoming, acessado em 04 de maio de 2021, no endereço: <http://weather.uwyo.edu/upperair/indices.html>.